경제학석사학위논문

Inference on the political network between Korean legislators

Evidence from the 20th Korean National Assembly

대한민국 국회의원들 간 정치적 네트워크에 대한 추론 : 20대 국회에서의 실증적 증거

2019년 8월

서울대학교 대학원

경제학부 경제학 전공

김 상 범

Abstract

Inference on the political network between Korean legislators

Evidence from the 20th Korean National Assembly

Sangbeom Kim
Department of Economics
The Graduate School
Seoul National University

This paper analyzes the political network between Korean legislators using bills cosponsored by legislators during the early 20th Korean National Assembly. This paper focuses on detecting hidden communities underlying the co-sponsorship network via non-negative matrix factorization concept. It is shown that the overall picture of the hidden political alliance is described as a confrontation between the Saenuri party, the ruling party of the period, and the opposition parties. This phenomenon is also shown by the type-specific community detection. In addition, by classifying the whole period in to three specific periods, we find that the unity of the Saenuri party decreases over time and this phenomenon is more manifest in within-party community structure. From these results, it is possible to guess that the decrease in unity may come from specific events of the period such as the big scandal related to former president and former leader of Saenuri party.

Keyword : Social network, Co-sponsorship network, Overlapping community detection, non-negative matrix factorization
Student Number : 2015-22505

Table of Contents

Tables

Figures

Chapter 1. Introduction

Over the last decade there have been many studies about networks in various fields such as bio-medical, physics, and computer science. By understanding the network in which a large number of objects are linked to one another, it is possible to understand why a serious disease such as cancer occurs or how a neuron responds to different kinds of stimuli.

In social sciences, studies on networks usually have been conducted under the name of "social network analysis". Network analysis is the process of examining the social structure via concepts of network and graph theory, characterized in terms of nodes and the edges connecting them. From this structure, one can infer how nodes or individuals are linked, the strength of those links, or which nodes are important in forming a network.

Social network analysis has been conducted on different types of social connections such as social media network, friendship network, or kinship network in a family. In this paper, the political network between Korean legislators is analyzed by the number of bills jointly proposed by two or more legislators. This is an example of a 'co-sponsorship network'. Studies on the topic mainly focus on finding important nodes or edges using various statistics such as centrality, connectedness, mutuality, and so on in order to fully describe co-sponsorship networks.

Even though just describing a network which has already been formed is important, other research that centers on network formation has been conducted. The random network model, which assumes that the link between pair-wise node is determined probabilistically and this probability depends on certain explanatory variables, is often used for this purpose. The ultimate goal of this probabilistic model is to figure out what factors are significant in forming a network. Among the various mechanisms that form the co-sponsorship network between legislators, party membership is cited as the most critical factor. When it comes to explaining political alliances, party membership alone cannot fully explain the interactions that occur between individual legislators across different parties, the mechanisms of which can be both varied and complicated.

In this paper, the main point is to figure out hidden communities via overlapping community detection algorithm to capture the political alliances more precisely. Even under a two−party system, several political factions could exist, temporary or otherwise, which can effectively make it into a multi−party system. Conversely, under the multi−party system, if a serious conflict between the ruling party and the opposition can turn it into a two−party system. These possibilities highlight the need to more precisely capture the political alliances through the overlapping community detection.

The fact that this paper centers on analyzing bills proposed during the 20th Korean national assembly has another important meaning. The reason why we choose this period as a subject of analysis is that during the 20th national assembly there have been several important political events that can lead to changes in political network or relationships between parties or legislators. Those events caused a chaos that led to changes in legislators' political behavior and if we can use more data about the bills proposed afterwards, some interesting analyses can be performed to see whether the changes are temporary or more persistent.

This paper shows that while there are 3 hidden communities in Korean congress, only 2 of those are politically important. The first is the conservative community represented by the Saenuri party, while the second is the liberal community represented by the Democratic party, the Justice party, and the People's party. In other words, the analysis indicates that there is a confrontation between the Saenuri party and the others.

Moreover, this paper conducts more detailed analyses by considering different types of bills and time periods. This is based on the idea that the shapes of political alliances can be dependent on the types of bills or differ across time. More specifically, if there is a political shock, the involved party shows relatively lower unity around its occurrence. If this phenomenon lasts for a long period of time, one may suppose that other political events can occur afterwards. Therefore, this paper is significant in that it allows us to capture the political relationships that are not revealed at the party level and infer the possibility of further political events based on the changes in community structure.

The rest of this paper is organized as follow. First, chapter 2 reviews the literature related to this paper, then chapter 3 explains the data and presents the descriptive statistics. After that, chapter 4 explains the estimation model and the result for the community detection. Finally, chapter 5 concludes and summarizes the main thrust of this study.

Chapter 2. Literature review

2.1 Co-sponsorship network analysis

There have been several studies to analyze a co-sponsorship network. Most of those focus on describing the characteristics or shapes of co-sponsorship networks and trying to determine what factors are important for the formation of such networks.

Fowler(2006) shows the most central legislators and pair-wise legislators who are strongly connected through the various concepts such as centrality and connectedness. Centrality is represented by various measures such as degree centrality, closeness centrality, betweenness centrality, and eigenvector centrality. Degree centrality essentially says that a node's importance is proportional to the number of edges it receives. Closeness centrality, on the other hand, is based on the idea that a more prominent node should be able to reach others through a smaller number of nodes. Betweenness centrality measures a node's importance by looking at how often it must be passed through when traveling between different points in the network. Finally, the eigenvector centrality is a measure that identifies how many links a node has with the others. The study also introduces the concept of 'legislative connectedness' to show how powerfully linked the legislators are to one another. The goal is to depict the network that exists in U.S. congress through these measures.

Chang (2011) aims to understand Korean congressional politics by analyzing the co-sponsorship network within. Based on the centrality and connectedness measures as used in Fowler(2006), the study tries to understand the characteristics of the co-sponsorship network in the 17th Korean national assembly. It examines what parties or legislators are important in forming the co-sponsorship network. Initially, it shows the overall picture of the co-sponsorship network, then it defines multiple time periods and identifies the characteristics of the network in each period.

As stated above, both Fowler(2006) and Chang(2011) try to describe the co−sponsorship networks that exist in a congress through certain measures. However, these are not sufficient to explain how different attributes of nodes or specific network patterns can cause the formation of a co−sponsorship network. Whether the network shows the significant transitive relationships between nodes or how significantly certain individual characteristics of nodes affect network formation cannot be captured just by describing the network. The exponential random graph model(ERGM) was developed to identify the influential factors in forming a network. In the model, the number of links are probabilistically determined by the attributes of nodes or specific patterns of network. It assumes that the number of links follows a probability model in the exponential family. In the particular case of describing co−sponsorship networks in a congress, the number of links between pair−wise legislators is count data, thus the Poisson distribution is assumed.

Park and Jang(2017) is an example of a study that uses ERGM to this end, using the data on the bills proposed during 17th national assembly. It tries to figure out what network statistics and covariates have a significant effect on the formation of the co−sponsorship network there. It considers mutuality, party membership, and number of times each legislator is elected as potential factors that form the co−sponsorship network, which are confirmed to be mostly significant. In particular, the most important factor is found to be party membership. The importance of parties in the legislator' s tendency to jointly propose bills has been confirmed through numerous studies, which is also evident from the actual political alliances in the parliament.

2.2 Community detection

While party−membership is found to play an important role in forming co-sponsorship networks [Park and Jang(2017)], it still cannot fully explain the political relationship between legislators. There are more than two parties in Korea, which can lead to a complicated network between legislators or between parties. In particular, minority parties have relatively more incentive to unite with the major parties because they cannot achieve their political goals without forming political alliances. According to the 'National Assembly Law', at least 10 legislators are necessary to propose a bill. This means that certain parties, such as the Justice Party which only has 6 members, cannot propose a bill without forming an alliance. This constraint also means that the political network in Korean politics can become quite complicated. Moreover, political factions within each party can cause a network to become more divided, which may lead to a decline in cohesion. This shows the importance of finding an explanatory framework that is more precise to fully understand these complex phenomena. Detecting hidden communities within the network can allow us to do that.

The term 'community' may be confused with a simple group or a political party in our context. However, community is something that should be found and not something that is a priori known. The existence of a community in a network is assumed, and can be detected through different approaches. Although it can be defined differently in various contexts, we define community in this study as subgraphs in which the nodes within are connected much more closely by those without[Fortunato (2010)]. This is similar to the basic idea of clustering; to cluster the similar things together and separated those that are different based on low within-group variance and high between-group variance.

There are two main views on communities. One assumes that communities are non-overlapping, while the other characterizes them as overlapping. Of these, the more traditional perspective was to consider communities as non-overlapping. Kernighan and Lin(1970), which proposed a method for partitioning arbitrary graphs into subsets, is one example of a study that takes this stance. Another such example, Potehn et al.(1990), presented a spectral partitioning algorithm that finds a vertex separator of a graph. The common feature of these studies is that they consider only a bisection of networks. In other words, these studies focused on simply partitioning a network into two subgraphs.

In the real world, however, there may be more than two communities in a network. Girvan and Newman(2002) proposed a new method for community detection that was also more realistic. The paper, using an unweighted network, describes the community structure of the graph by removing edges with high edge betweenness, based on the idea that this is indicative of those connecting communities. Aaron et al.(2004) proposed a generalized version of this algorithm for weighted networks.

In the real world, however, a node can also interact with those in different groups, which implies that it is more natural to assume that communities can overlap. If we were to take non-overlapping community detection algorithm, the results would be indistinguishable from simply identifying the existing parties. In this paper, however, the aim is to analyze the political alliances between parties more accurately using the assumption that communities can in fact be overlapping.

Palla et al.(2005) supports this view, based on the observation that a network is usually large in scale. It introduced an approach for analyzing communities based on this assumption, leading to the development of new methodologies. Among these, we applied non-negative matrix factorization as proposed by Psorakis et al.(2011), to our real world data set on co-sponsored bills by legislators of the Korean National Assembly. The concept of matrix factorization is particularly helpful in determining the existence of latent communities and each legislator's degree of membership in each community.

Chapter 3. Data

This study uses cosponsored bills by legislators to detect hidden communities underlying the co-sponsorship network in the 20th Korean National Assembly. About three thousand bills have been jointly proposed during the early days of the 20^{th} Korean National Assembly, specifically from June 2016 to November 2016. The following subsection shows that the willingness to participate in cosponsoring differs across parties, then goes on to show how the adjacency matrix describing the network to be used in our estimation is constructed from our dataset.

3.1 Parliamentary seat

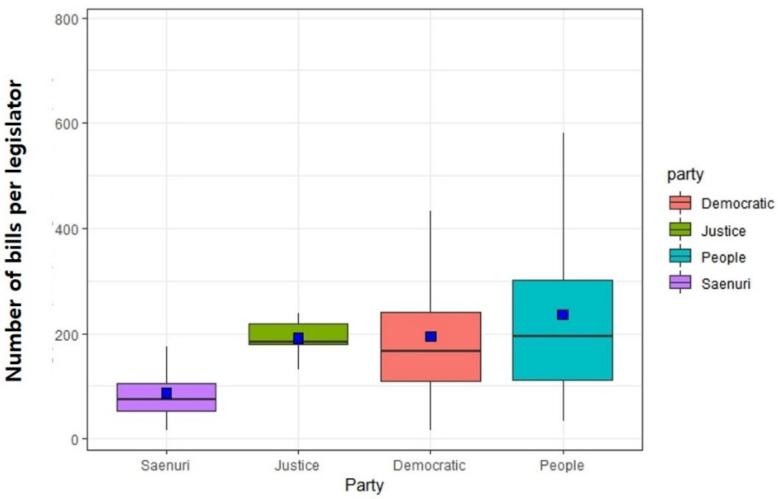
The distribution of Parliamentary seats is a rough representation of the political power structure, showing which parties are politically powerful and which are less so.

Table 1 Parliamentary seat

Party	No. Legislators	Party	No. Legislators
Saenuri Party	122	People's Party	41
Democratic Party	123	Justice Party	6

Table 1 shows that the Saenuri Party and the Democratic Party have the most seats in the National Assembly, and therefore are the two leading parties. In the early 20^{th} National Assembly, while the Saenuri party was the ruling party and the Democratic party was the main opposition party, neither had a distinct advantage over the other. This was the first time such a situation had occurred since 2000 and it was expected that this could lead to a complicated political geography and unexpected events. Figure 2 shows the average level of participation in cosponsoring bills in each party.

Figure 1 Box plot of number of bills per legislator



On average, legislators in the People's Party tended to propose many more bills than the two major parties and the same could be said for the Justice Party as well. This may have been because both parties are relatively minor, leading to each legislator in the two parties being more eager to participate in the proposal of bills compared to those in the two leading parties. This is a natural phenomenon in terms of getting political power. Although the Democratic Party was one of the leading parties during this period, it was not the ruling party. Its role was to keep the Saenuri party in check, the ruling party at the time, and to accomplish this they had to form alliances with members of other parties. As a result, figure 2 shows that there are strong incentives to propose bills jointly for both minor and major parties.

3.2 Adjacency matrix

Legislators not only vote on whether to pass bills but also propose bills to express his or her opinion. Moreover, this is usually done in concert with other legislators. We assume that these jointly submitted bills reflect the level of interaction among legislators. We therefore have represented these scenarios as a matrix, also called “adjacency matrix”. A typical adjacency matrix for an “unweighted” network is defined as follows :

$$\mathrm{A_{ij}} = \begin{cases} 1 \ \ \textit{if there is a path from node i to j} \\ 0 \ \ \textit{otherwise} \end{cases}$$

This definition depicts a basic adjacency matrix of networks, but it only shows whether interaction exists between two nodes. Examples include whether someone follows my Twitter or not, or whether someone is my friend or not. However, in this study we consider the number of bills jointly submitted as the intensity of interaction, and thus the elements of the presented adjacency matrix is not binary. Accordingly, the adjacency matrix used in our estimation is defined as follows :

$$\mathrm{A_{ij}} = \begin{cases} \alpha \ \ \textit{number of bills jointly submitted by legislators i and j} \\ 0 \qquad\qquad \textit{no joint submission} \end{cases}$$

It is important to note that the network of interest in this analysis is undirected. In an undirected network, the link between two nodes does not have a specific direction, which means the network is symmetric. Of course, co-sponsorship networks can also be expressed as directed networks. This is the case with James H. Fowler(2006) and Park and Jan(2017), where ties between legislators i and j reflect the support of one for the other, the direction of which can be from i to j or vice versa. In this case, symmetricity of the adjacency matrix cannot be guaranteed.

However, the assumption that the co-sponsorship network is directed may be too strong in the context of the Korean national assembly. This would have been reasonable if legislations are proposed purely out of the legislators' political will and sincere efforts are made to try and persuade other legislators into cosponsoring bills. However, this is rather unlikely in this particular context, making the assumption quite inappropriate.

For this reason, the network is instead assumed to be undirected in this analysis. Fortunately, there have also been several studies that assumes undirected co-sponsorship network (Mirko Signorelli et al 2017; Christian S. Schmid et al 2017). In addition, α is a non-negative integer, which indicates that we are assuming an weighted network. Since the number of legislators is 300, we have constructed a 300 by 300 symmetric adjacency matrix.

Chapter 4. Model and Empirical results

4.1 Model

This study applies overlapping community detection by non-negative matrix factorization (Psorakis et al. (2011)). Initially, matrix factorization was a linear algebraic tool to efficiently solve linear equations, but it has since been used in various fields such as education, psychology and business to effectively realize their own specific purposes. In particular, the tool is used in business for a recommendation algorithm system that decomposes a rating matrix into two lower rank matrices.

In our context, this is done by decomposing the adjacency matrix A into two lower rank matrices W and H. This can be expressed mathematically as follows :

$$\mathrm{W},\mathrm{H} = \underset{W,H}{argmin}\|A - W \cdot H\|^2 \qquad (1)$$
$$\text{where } \mathrm{A} \in \mathrm{R}^{\mathrm{n}\times\mathrm{n}}, W \in \mathrm{R}^{\mathrm{n}\times\mathrm{K}}, H \in \mathrm{R}^{\mathrm{K}\times\mathrm{n}}$$

Since the adjacency matrix A is symmetric, H is equal to W' and hence, equation (1) is equivalent to

$$\mathrm{W} = \underset{W}{argmin}\|A - W \cdot W'\|^2 \qquad (2)$$

However, there are three fundamental issues in detecting communities by non-negative matrix factorization that need to be addressed. The first is how to select K, the rank of matrix W. After leaving out individual components or full rows of the adjacency matrix, we could implement matrix factorization with the remaining values repeatedly by changing the size of K. Then we can get different W depending on the size of K, construct $\mathrm{W}\cdot\mathrm{W}'$ which is the predicted value of A, and calculate the prediction error from the difference between the actual values that were removed from the original adjacency matrix and the corresponding predicted values. We could thus find the optimal K that minimizes this.

The second issue is how to identify the community structure from the lower rank matrix W. This is closely related to how to interpret the elements of W. Basically, $k \in K$ is the unknown and latent dimension of W. It is hard to explain the meaning of dimension k explicitly. This is usually explained a posteriori and the element w_{ik} refers to how i is assigned to dimension k. If we define k as a community, w_{ik} indicates membership of i to community k. Hence, from w_{ik}, we can identify the community structure.

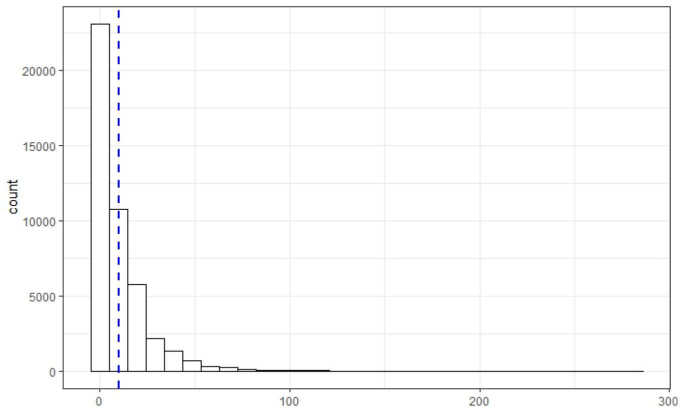
The third issue is that the adjacency matrix is very sparse as shown in figure 1. According to figure 1, this adjacency matrix is subject to high sparsity and over-dispersion. This is shown by the fact that more than 10,000 elements are zeros while the maximum value for A_{ij} is 281. In addition, its mean and standard deviation is 10.26 and 16.27 respectively. This sparsity causes uncertainty, and hence, factorizing the matrix as stated in equation (2) may generate inaccurate estimation results. To handle this drawback, Psorakis et al.(2011) used the probabilistic approach for non-negative matrix factorization. The probabilistic model is described below :

$$\mathrm{p}(\mathrm{A}|\mathrm{W}) = \mathrm{p}\left(\mathrm{A}\middle|\hat{A}\right), where\ \ A{\sim}poisson\left(\hat{A}\right) \tag{3}$$

$$\mathrm{w_{ik}} \sim Half\ \ normal\left(0, \beta_k^{-1}\right) \tag{4}$$

$$\mathrm{\beta_k}{\sim}gamma(a, b) \tag{5}$$

Figure 2 Histogram of the number of bills proposed by pairs of legislators



Equation (3) indicates the likelihood of observation. It is assumed that the number of bills that legislator i and j jointly propose[①]) follows a Poisson distribution with mean $\hat{A} = W \cdot W'$. According to equation (4), w_{ik} has half normal distribution as a prior distribution, whose mean is equal to zero and precision is β_k. The assumption of the half normal distribution can be attributed to the definition of non-negative matrix factorization where the elements of two matrices decomposed from A should be positive.

β_k is assumed to have a gamma prior with fixed parameters a and b. The β_k is a parameter that controls the strength or importance of the latent community k. If β_k becomes larger, then w_{ik} is centered around zero, implying that w_{ik}^2 is also close to 0. This means that community k cannot explain the level of interaction between node i and j.

On the basis of these distributions, we can set a posterior distribution as stated below. We want to maximize the posterior

① Note that during the actual estimation process, we used normalized adjacency matrix where all elements lie between 0 and 1.

distribution with respect to β and W, to get point estimates of β and W. This means that this whole process is not fully Bayesian because we do not aim to get probability distribution of each parameter.

$$p(W, \beta|A) = \frac{p(A|W)p(W|\beta)p(\beta)}{p(A)} \tag{6}$$

$$W, \beta = arg\max_{W,\beta} p(W, \beta|A) \tag{7}$$

In the next sub section, we present the empirical findings resulting from the above estimation process.

4.2 Overlapping community detection result

By applying overlapping community detection, we can calculate the degree of membership of each legislator in each latent community. This approach can provide more informative results than the simple clustering method, which assumes that each node belongs to only one community.

As a measure of membership, w_{ik}, legislator i's level of membership to community k could be an option. However, this value could be affected by the absolute level of participation in bill proposals. In other words, if a legislator does not participate in the proposal of bills at a high rate, the level of membership to each community can still be low. How frequently a legislator participates in proposals depends on various individual characteristics or situational factors, so it is necessary to handle this problem. To address this situation, an alternative measure called soft membership, is used in this paper. It is defined as follows :

$$s_{ik} = \frac{w_{ik}}{\sum_k w_{ik}} \tag{8}$$

Equation (8) shows that soft membership aims to capture the membership to community **k** of legislator **i**. Even if legislator **i**'s absolute value of membership in each community is small, its soft membership to community k can be larger or smaller relative to others.

4.2.1 Community structure - Base results

The result based on this soft membership is shown in figure 4.

Figure 3 Overlapping community detection result

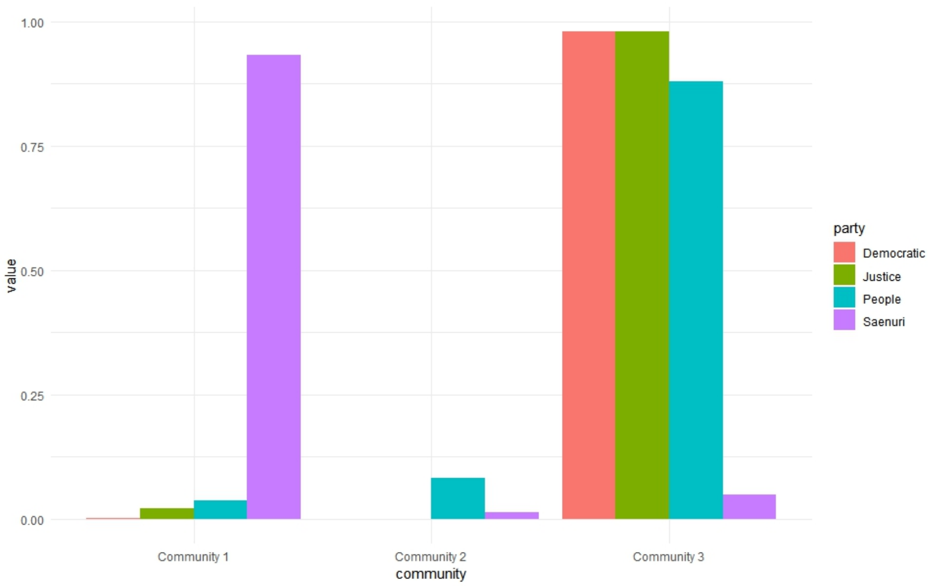
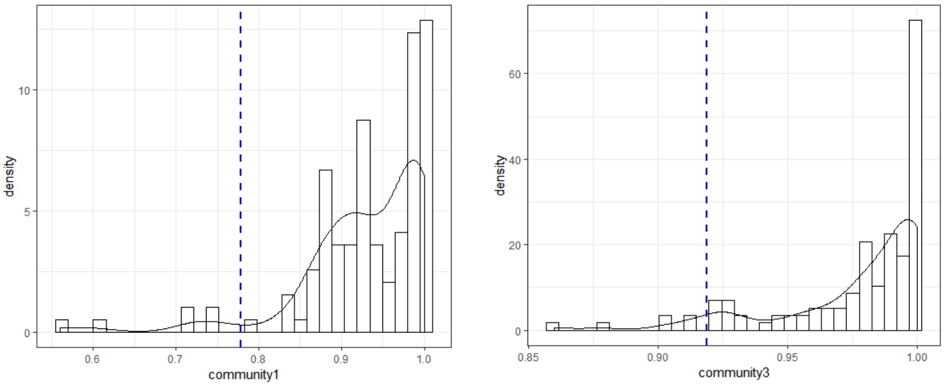


Figure 4 shows that 3 communities are identified by the estimation. The Democratic Party and the Saenuri Party very likely dominate community 3 and community 1, respectively. In Korea, the Democratic Party represents the liberal side while the Saenuri Party represents the conservative side. Figure 4 also shows that members of Justice Party show a high level of soft membership to community 3 on average. The Justice party has been regarded an extremely liberal party, and it tends to collaborate with the Democratic Party. Legislators in People's party are mostly assigned to community 3, thus we can infer that there exists a severe confrontation between a ruling party and opposition parties during the sample period.

Figure 4 Histogram of soft membership



Panel (a) Histogram of soft membership to community 1 (Saenuri Party)

Panel (b) Histogram of soft membership to community 3 (Democratic Party)

From the result shown in figure 3, we can define community 1 as the main community for the Saenuri party, in that the legislators in the party shows relatively higher degree of membership to that community relative to others. Similarly, community 3 can be defined as the main community for the Democratic party and while the remaining parties can be thought of as having their own main community.

The fact that legislators in minor parties have their membership spread evenly across various communities is not so surprising given that this allows them to achieve their goals more easily. However, if those in major parties are shown to have a stronger membership in communities other than his or her parties' main community, a different interpretation can emerge. This situation can in fact represent a "deviation" from the main community which may be an indicator of divisiveness in the party. In other words, members of major parties may be colluding with those of other parties over their colleagues, in which case the deviations indicate a serious situation for the party. To identify these deviators, we set a cut-off point and find those who are below this point.

Figure 4 shows the distribution of legislators in the Saenuri party and the Democratic party assigned to their main community. Based on this distribution, the cut-off point is calculated by the mean and standard deviation of the participation ratio(i.e, $\mu - 2\sigma$). Legislators that show a level of soft membership below the cut-off point can be seen as being unusual compared with other legislators in their parties. Table 2 reveals the level of deviation from the main community for the Democratic Party and the Saenuri Party.

Most of the members in panel (a) have community 3 as their second community. Recall that community 3 is the liberal community, ideologically different from community 1. Their relatively lower soft membership to community 1 and higher soft membership to community 3 can be seen as being representative of a fissure within Saenuri party. Panel (b) shows the name of legislators in the Democratic party, who show relatively lower soft membership to its main community. Although the legislators are unusual compared with other legislators, all legislators in Democratic party shows absolutely high soft membership to its main community and hence, there is no evidence that shows a fissure within the party.

Even though there are several legislators with relatively lower membership to the main community, it is not sure that those deviated legislators directly reveals that there is a factional situation in the Saenuri party. That is still a matter of debate because from the overall picture of political alliance, the Saenuri party stand alone in community 1, meaning that there is no fissure within the party. We can reasonably question that the overall picture can mask the factional situation that can occur around the presidential impeachment and this will be shown in the period specific community detection.

Table 2 Members with relatively low soft membership to main community

Name	Community 1	Community (2nd high membership)
Yoo, Seungmin	0.56	Community 3 (0.43)
Yoon, Sanghyun	0.60	Community 2 (0.29)
Hong, Moonpyo	0.72	Community 3 (0.25)
Kang, Gilboo	0.75	Community 3 (0.23)
Ahn, Sangsoo	0.75	Community 3 (0.23)

Panel (a) Members in Saenuri party that are more likely to deviate from the main community

Name	Community 3	Community (2nd high membership)
Min, Hongcheol	0.86	Community 1 (0.14)
Jung, Sungho	0.88	Community 1 (0.12)
Hong, Euirak	0.90	Community 1 (0.10)
Oh, Jese	0.91	Community 1 (0.09)
Park, Beomgye	0.91	Community 1 (0.09)
Lee, Gaeho	0.91	Community 1 (0.08)

Panel (b) Members in Democratic party that are more likely to deviate from the main community

4.2.2 Community structure with respect to the types of issues

The previous section shows that community structure represents a confrontation between the ruling party and the opposition parties. However, one needs to consider the fact that the estimation does not take into account differences across specific issues or time periods. In this section, therefore, the community structure is estimated with the types of bills proposed by legislators taken into account.

The reason why specifying the types of bills can be important is that it is hard to define a political party's ideological orientation on one dimensional space. Their ideologies are generally multifaceted, so there may be different shapes of alliances between parties, which can lead to different forms of community structure.

Park(2008) notes that there are competitive structures between political parties in national assembly in the 17th national assembly. Since each political party has its own political agenda and different points of emphasis, the competitive structure can change depending on the types of issues in question. According to the study, there was a conflict over inter-Korean relations between the liberal parties and the conservative party[②], over market regulation, welfare expansion, and so on. Jang(2017) shows that in the 18th and 19th national assembly, there were confrontations over the bills on migration policy between the ruling party and the opposition parties.

From both studies, one can infer that historically in Korea, the most important part of the competitive structure is that of the liberal side versus conservative side and that of the ruling party versus opposition parties. In various fields, each party has different agendas so that there is a possibility that unexpected alliance can be made and that may depends on what types of bills the legislators in each party tries to propose.

To identify the bill-specific community structure, this paper

② In 17th Korean national assembly, on the liberal side, there were the 'Our party' and the 'Korean Democratic Labor Party', and on the conservative side, there was the 'Grand National Party' classify all bills by three categories: economy, welfare, and education. Those three categories are considered the most important topics for people and categorizing is based on the types of bills and committees because each bill is in control of corresponding committee.

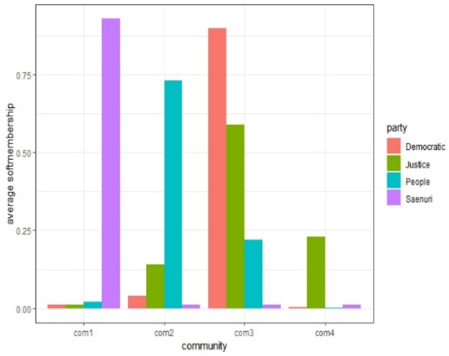
The number of bills in each category is shown in table 3.

Table 3 Number of bills per legislator by party and category

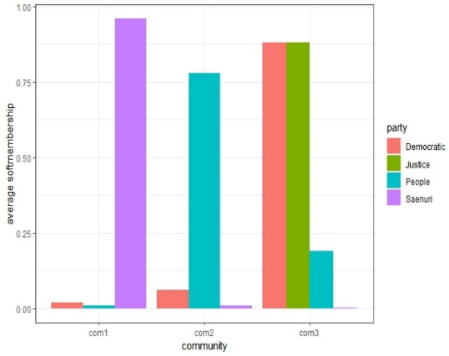
Category	Number of bills	Number of bills per legislator
Economy	667	
	Saenuri Party	15.34
	Democratic Party	38.93
	People's Party	52.95
	Justice Party	37.67
Health and welfare	236	
	Saenuri Party	6.91
	Democratic Party	14.23
	People's Party	15.29
	Justice Party	19.83
Education	152	
	Saenuri Party	3.70
	Democratic Party	9.75
	People's Party	10.20
	Justice Party	11.83

According to the table 3, legislators in People's party and those in Justice party cosponsors bills more frequently than those in other parties. This is the similar phenomenon shown in figure 1. Category “Economy” includes bills which are related to tax, insurance, labor and so on.

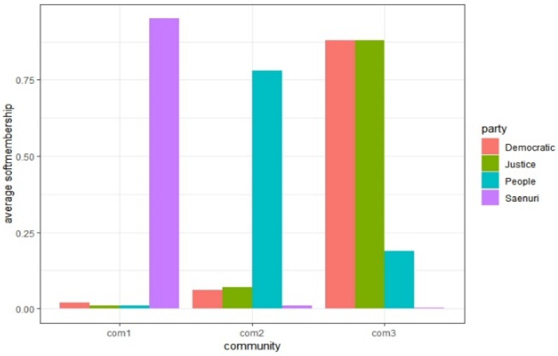
Figure 5 Type−specific community structure



Panel (A) “Economics”



Panel (B) “Welfare”



Panel (C) “Education”

Panel (A) in figure 5 shows the community structure estimated using the “Economy” bills. It indicates that Saenuri party does not tend to collaborate with other parties, People’s party is highly likely to propose bills independently but it also try to collaborate with Democratic and Justice party. Those three parties form an alliance in community 3. They mostly propose bills which are related to income redistribution, increasing corporate tax, increasing minimum wage, and so on. These bills are quite contradictory to the political direction of Saenuri party and this difference in political direction may cause it to be isolated from other parties.

Category “Health and welfare” includes bills related to people’s basic human rights, hygiene, welfare facilities and so on. Panel (B) shows the community structure estimated using the “Health and welfare” bills, similar to the results shown in panel (A). Legislators in each party proposes bills about how to support people; providing job education service for the young who are seeking jobs, guaranteeing health-care service for the old, and subsidy for child support.

Even though legislators in Saenuri party and those in other parties propose bills with similar purposes, Saenuri party does not tend to collaborate with other parties, and other 3 parties form an alliance for proposal. This phenomenon occurs also in “Education” and hence, it can be conjectured that ruling party was isolated and other opposition parties including Democratic, People, and Justice party form a strong alliance and check Saenuri party in the early 20th national assembly.

4.2.3 Period specific community structures

So far, the estimation of the community structure has been done without specifying the time period. However, it is obvious that political relationship can change over time due to some important shocks, political or otherwise. This section tries to capture this change in community structure, which leads to a better understanding of the different aspects of political alliances between parties.

To make the specification of periods more valid, we define three specific time periods based on politically important events in the

Korean national assembly. The three time periods defined in this paper is as follows :

1) Period 1 : Beginning of the 20th Assembly
2) Period 2 : Conflict on THAAD
3) Period 3 : Public revelation of Soonsil Choi's scandal

As a result of the general election, the Saenuri party lost its position as a dominant party in the national assembly. Instead, the Democratic party occupied the largest number of parliamentary seats, which gave it an opportunity to lead the national assembly. The main reason of the Saenuri party's failure in the general election was the series of conflicts between the factions within. At the beginning of the 20th assembly(from the end of May to July), the leader of the Saenuri party pledged to reform the party to address this factionalism. Leaders of other parties followed suit, pledging to avoid infighting within the party and to focus more on problems concerning public welfare. This political atmosphere suggests that the unity of each party would have been strong.

During period 2(from August to September), there was a severe conflict between the Saenuri party and the Democratic party regarding the Park administration's decision to install THAAD in Korea. In this period, therefore, there may have been a high degree of political division in the national assembly.

After Soonsil Choi's scandal was publicly released, the political situation changed dramatically. Legislators in opposition parties criticized President Geun Hye Park and the Saenuri party and tried to change the government. Within the Saenuri party, there have traditionally been two factions: the 'Pro-Park' and the 'Anti-Park'.

During period 3, there would have been a severe conflict between these two factions, leading to a complicated political situation both across parties and within the Saenuri party.

Based on these periods classified by important political events, we implemented the identical estimation process.

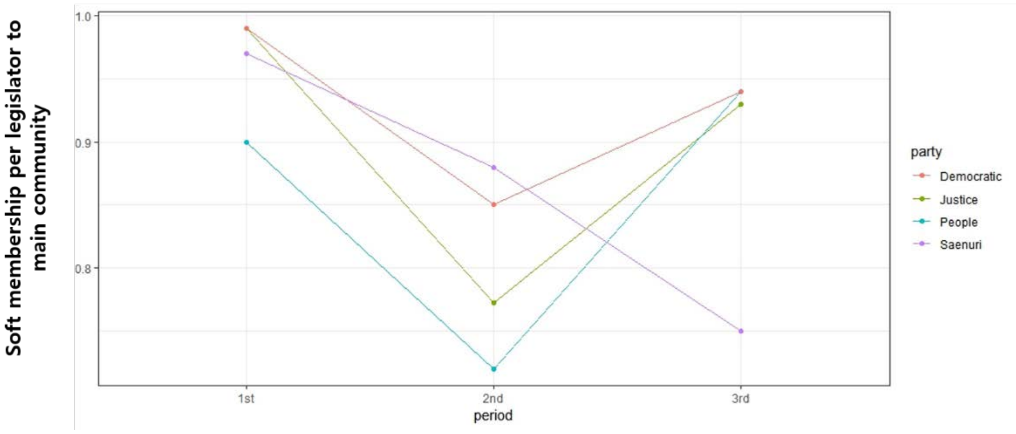
Table 4 Number of bills per legislator for each period

	Period 1	Period 2	Period 3
Total number of bills	1,135	1,162	970
Number of bills per legislator			
Saenuri	33.36 (15.55)	29.2 (18.6)	24.6 (16.7)
Democratic	79.9 (48.63)	66.88 (47.50)	50.01 (37.26)
People	80.26 (67.18)	87.85 (61.20)	70.27 (50.32)
Justice	59.3 (9.5)	72 (17.76)	60.83 (16.36)

Figure 6 shows the change in soft membership of legislators to the main community of each party. In period 1, all parties show high soft membership to their respective main community. Although there are declines in soft membership in period 2, this is common to all parties and can be interpreted in various ways. There may have been active cooperation between parties, or this may simply be a temporary phenomenon. However, it is hard to explicitly interpret this phenomenon during period 2.

In period 3, the average soft membership of the Democratic party, the Justice party and the People's party to their main communities reverted to higher levels. On the other hand, the Saenuri party's average soft membership of legislators decreases further. One can thus infer that this decline in soft membership may come from the characteristic of period 3. This is the time when there was a big political scandal related to former President Park and one can suppose that the 'Anti-Park' faction tried to distinguish itself from the 'Pro-Park' faction to show that they are unrelated to the scandal. In this sense, this decline is somewhat expected and this may represent a deterioration of unity within the party.

Figure 6 Change in average level of soft membership to the main community of each party

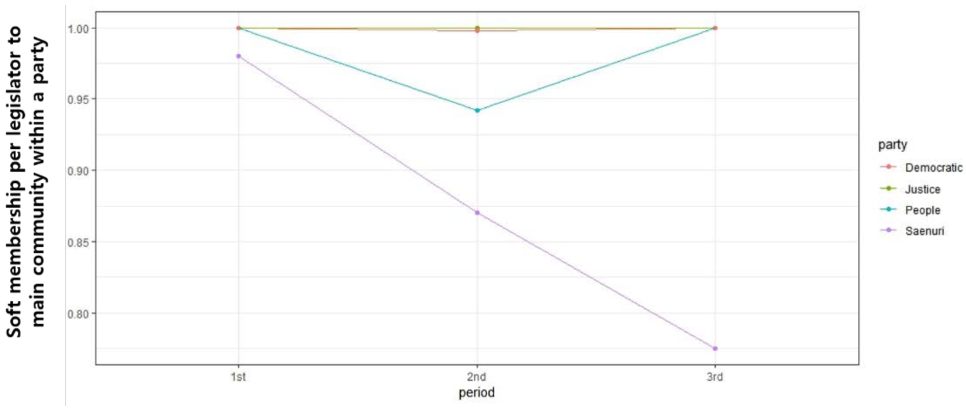


4.2.4 Period specific within-party community structures

From the changes in average soft membership of legislators to the main community, it is suspected that there is a decline in the unity of the Saenuri party. To show this more clearly, this section tries to find community structure within a party. Political factions, both established and temporary, can exist within a party and the shapes of the factions can differ across time. By looking at this within−party community structure, it is expected that we can more precisely capture the existence of factionalism and the level of unity within each party.

For this process, it is necessary to reconstruct the adjacency matrix. Let A_t^p be the adjacency matrix constructed with only pair−wise relationships between legislators in the same party '**p**' in period '**t**'. We implement the same estimation process using A_t^p to capture the community structure within a party. Figure 7 shows the change in the average soft−membership of the legislators to the main community within each party.

Figure 7 Change in average level of soft membership to main community within each party



The Democratic party, the Justice party, and the People's party show soft membership above 0.9 but the Saenuri party's soft membership constantly decreases from period 1 to period 3. This allows us to infer that the unity of the Saenuri party decreases because of the big scandal.

Chapter 5. Conclusion

While there are many possible approaches to analyze co-sponsorship networks, this paper focuses on detecting hidden communities underlying the co-sponsorship network. There is no denying that describing a co-sponsorship network using network statistics such as centrality and connectedness, and figuring out what factors are significant in network formation by exponential random graph model(ERGM) can give valuable information. In terms of capturing hidden relationship between parties however, community detection can be more useful because it allows us to find out how political parties form alliances more precisely than the other two approaches.

As in Psorakis et al. (2011), it is assumed that underlying communities are overlapping, so each legislator can be assigned to more than one group. Our initial analysis detects 3 communities statistically but of these, only 2 communities are actually important in the co-sponsorship network, suggesting a confrontation between the ruling party and opposition parties.

The results are similar even if we take the various types of bills into account when detecting community structures. No matter what the types of bills are, the form of political alliances in the co-sponsorship network can still be characterized as the confrontation of the ruling party versus opposition parties. The analysis is based on the expectation that the forms of political alliances can depend on what types of bills legislators propose because each party's political orientation is multifaceted. A party can have an extremely liberal attitude on the 'economics' side but have a conservative attitude on other types of bills. The results, however, show that the types of bills are not important. The Saenuri party is still found to have been isolated from other parties, and it could not lead the national assembly despite being a ruling party at that time.

Moreover, the existence of legislators in the Saenuri party who have an unusually low level of soft membership to its main community shows another side of the political alliances. This can be interpreted as a 'deviation' from the main community, which is an important phenomenon in that the impact of this 'deviation' in a major party would be more critical than that in minor parties.

The 'deviation' is related to the unity of the party and to capture this, we try to detect the community structure by dividing the sample period into 3 separate ones based on political events, and within-party community structures in each period. According to the result, the average soft membership of legislators in the Saenuri party to its main community decreases over time, and this decline in unity can be thought to implicitly reveal the existence of factionalism in the party.

We may further suppose that this phenomenon arises due to events specific to the period, the big scandal of the former president. This would also imply that the barriers to the impeachment of the former president may have been much lower than was originally expected. In addition, one could even suppose that had this factionalism persisted, a splitting of the party could have occurred. These inferences come from the community detection results that account for various factors such as types of bills, time periods, or both. The quantitative results that we obtained from the estimation can also explain political phenomena that occur today, which makes this study more meaningful.

References

Chang, D. (2011). Policy Networks in the 17th National Assembly of Korea: An Analysis of Co-sponsorship and Friendship Ties. *Korean Party Studies Review, 10*(2), 157-187.

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E, 70*(6), 066111.

Fortunato, S. (2010). Community detection in graphs. *Physics reports, 486*(3-5), 75-174.

Fowler, J. H. (2006). Connecting the Congress: A study of cosponsorship networks. *Political Analysis, 14*(4), 456-487.

Girvan, M., & Newman, M. E. (2002). Community structure in social and biological networks. *Proceedings of the national academy of sciences, 99*(12), 7821-7826.

Jang, I. S. (2017). Legislative Cosponsorship Networks of Immigration Policy. *디아스포라연구, 11*(2), 7-50.

Kernighan, B. W., & Lin, S. (1970). An efficient heuristic procedure for partitioning graphs. *Bell system technical journal, 49*(2), 291-307.

Lazer, D. (2011). Networks in political science: Back to the future. *PS: Political Science & Politics, 44*(1), 61-68.

Mnih, A., & Salakhutdinov, R. R. (2008). *Probabilistic matrix factorization.* Paper presented at the Advances in neural information processing systems.

Newman, M. E. (2004). Analysis of weighted networks. *Physical Review E, 70*(5), 056131.

Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *nature, 435*(7043), 814.

Park, C., & Jang, W. (2017). Cosponsorship networks in the 17th National Assembly of Republic of Korea. *Korean Journal of Applied Statistics, 30*(3), 403−415.

Park, C. P. (2008). Party Competition Dimensions and Party Position. *한국정당학회보, 7*(2), 5−40.

Pothen, A., Simon, H. D., & Liou, K.−P. (1990). Partitioning sparse matrices with eigenvectors of graphs. *SIAM journal on matrix analysis and applications, 11*(3), 430−452.

Psorakis, I., Roberts, S., Ebden, M., & Sheldon, B. (2011). Overlapping community detection using bayesian non-negative matrix factorization. *Physical Review E, 83*(6), 066114.

Robins, G., Pattison, P., Kalish, Y., & Lusher, D. (2007). An introduction to exponential random graph (p*) models for social networks. *Social Networks, 29*(2), 173−191.

Schmid, C. S., & Desmarais, B. A. (2017). *Exponential random graph models with big networks: Maximum pseudolikelihood estimation and the parametric bootstrap.* Paper presented at the 2017 IEEE International Conference on Big Data (Big Data).

Signorelli, M., & Wit, E. C. (2018). A penalized inference approach to stochastic block modelling of community structure in the Italian Parliament. *Journal of the Royal Statistical Society: Series C (Applied Statistics), 67*(2), 355−369.

국문초록

본 연구는 20대 국회에서 공동 발의된 법안을 바탕으로 대한민국의 국회의원들 간에 존재하는 정치적 네트워크에 대한 실증 연구를 진행하였다. 음수 미포함 행렬 분해 방식을 이용하여 공동발의 네트워크에 숨겨져 있는 커뮤니티를 밝혀내는 데에 초점을 두고 있으며, 추정 결과 전체적인 정치적 지형의 형상은 당시 여당인 새누리당과 나머지 야당과의 대립구도를 보이고 있다. 이러한 모습은 발의된 법안의 속성에 따라 분류하여 추정을 한 경우에도 비슷하게 나타나고 있다. 게다가 보다 세부적인 분석을 위하여 정치적 사건을 기반으로 전체 기간을 세 개의 특수한 기간으로 분류하여 커뮤니티 구조를 밝히는 분석을 진행하였으며, 그 결과 새누리당의 결속력이 시간이 지남에 따라 떨어지는 것으로 나타났다. 이러한 현상은 발의 네트워크를 정당으로 한정하여 분석을 진행하였을 때 더욱더 명확히 나타났다. 본 연구에서는 이러한 결속력의 하락은 특정한 정치적 사건에 기인했을 것이라는 점, 나아가 결속력의 하락을 통해 대통령 탄핵이나 분당과 같은 정치적 현상들 대한 대략적인 추론을 정량적으로 할 수 있다는 점을 보여주고 있다.

Keyword : 소셜 네트워크, 공동발의 네트워크, 중첩 커뮤니티 검출, 음수 미포함 행렬 분해
Student Number : 2015-22505