

Statistical ERGM analysis for consulting company network data

Yejin Park^a, Jungmin Um^a, Subeen Hong^a, Yujin Han^a, Jaehee Kim^{1,a}

^aDepartment of Statistics, Duksung Women's University

Abstract

A company is a social group of many individuals that work together to obtain better results, and it is an organization that pursues common goals such as profit. As a result, forming networks among members, as well as individual communication abilities, is critical. The purpose of this research was to determine what factors influence the creation of employee advice relationships. Using the ERGM(Exponential Random Graph Model) approach, we looked at the network data of 44 individuals from consulting firms with offices in the United States and Europe. The significance of structural network factors like connectivity was first discovered. Second, the gender factor had the most significant main influence on the likelihood of adopting each other's advice. Third, geographical homogeneity resulted in higher link probabilities than major impacts of gender. This research looked at ways to make a company's network more efficient and active.

Keywords: company network, ERGM, network data, reciprocity, transitivity

1. 서론

현대 사회가 복잡해지고 발전함에 따라 의사소통 능력이 중요해지고 있다. 사회 구성원들이 맺는 사회적인 관계를 나타낸 사회적 네트워크(이하 네트워크)는 이제 우리 삶에서 불가결한 것이 되었다. 개인과 개인의 네트워크도 중요하지만, 조직 내에서도 네트워크의 영향력이 크게 작용하고 있다.

조직은 더 나은 의사결정을 위하여 조연 행동이 빈번히 일어나는 곳이다. 조직은 다양한 수준의 업무능력과 지식을 가진 개인으로 구성되어 있기 때문에 높은 성과를 창출하기 위해서는 개인 간 지식의 전이와 원활한 의사소통이 필수적이다 (Guk, 2012). 조직의 더 나은 의사결정은 더 큰 이익으로 이어지기 때문에 이에 관하여 어떠한 속성을 가진 직원이 유의미한 도움을 주고 있는지 파악하고자 한다.

개인이 맺고 있는 네트워크가 승진이나 임금 등과 같은 개인의 성과와 더불어 효율적인 의사결정, 생산성, 이직률 등과 같은 조직의 성과에 미치는 영향에 대한 관심이 증대되면서 이와 관련한 연구들이 이루어져 왔다 (Kang과 Park, 2018). Song (2017)은 기업 내에서 조연을 구하거나 하는 행동은 직원 간의 소통이라고 볼 수 있고, 이직률을 낮추거나 노사관계에 긍정적인 영향을 미치는 주요 요인인 것으로 파악하였으며 소통하는 기업이 그렇지 않은 기업보다 대체로 높은 기업성적을 보여준다고 하였다. 실제로 조연을 구하는 행위는 효과적으로 정보를 수집할 수 있고 인간적 문제 같은 리스크를 감수하면서도 조연을 구하는 것에 자신감을 보여줄 수 있다는 등의 긍정적인 효과가 부정적인 효과보다 더 많은 것으로 나타났다 (Brooks 등, 2015). 이처럼

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government(MSIT) (No. 2021R1A4A5028907) and Basic Research (No. 2021R1F1A1054968).

¹ Corresponding author: Department of Statistics, Duksung Women's University, 419 Samyang-ro 144 Gil 33, Dobong-Gu, Seoul 01369, Korea. E-mail: jaehee@duksung.ac.kr

비슷한 사람들 사이의 상호작용이나 접촉은 다른 사람들 사이의 접촉에 비해 훨씬 더 높은 비율로 발생하며, 서로 비슷한 개인들은 공통된 삶의 경험을 공유하기 때문에 관계를 쉽게 맺는다 (McPherson 등, 2001).

Rapoport와 Horvath (1961)는 연결 관계에서 약한 연결(weak tie)을 통해 전달되는 정보는 강한 연결(strong tie)을 통해 전해지는 것보다 궁극적으로 더 많은 수의 사람들에게 닿을 수 있다고 하였다 (Granovetter, 1974). 그러나 인적자원개발 차원에서 조직 내 다양한 네트워크 형성이 중요함에도 불구하고, Woehler (2017)는 남성은 네트워킹 행동으로 지위가 높은 친구를 가질 수 있는 것으로 연결되며, 여성은 네트워킹 행동으로 상대적으로 지위가 낮은 친구를 갖는 것과 관련 있다고 시사하였다. Stallings (2010)은 남성, 여성 모두 지위가 높은 사람을 선호하며, 여성의 직업 관계가 남성 여성 모두에게 선호도가 낮다는 등 일부 연구 결과가 있지만, 직장 네트워크 관련된 연구가 지속적으로 필요하다. 따라서 본 연구는 컨설팅 회사의 데이터(Cross와 Parker, 2004)를 수집하여 직원 간 조언 행동 네트워크 구조를 살펴보고, 지수족 랜덤 그래프 모형(ERGM)을 활용하여 조언 제공자와 조언 수용자의 속성을 중심으로 직장 네트워크의 조언 관계 형성에 어떠한 영향을 미치는지 고려하고자 한다. 따라서 본 연구는 다음과 같은 질문에 대한 분석을 하고자 한다.

(i) 어떠한 직원이 조언 관계에 있어 중요한 역할을 하고 있는가? (ii) 직원의 어떠한 속성이 서로 조언을 구하는데 가장 영향력을 미치는가? (iii) 조언 관계에서 성별, 직급, 지사의 위치가 동질성 혹은 이질성을 보이는가?

본 논문의 2장에서 ERGM을 소개하며 그 구성에 대하여 설명한다. 3장에서는 직장 네트워크의 기초통계량과 모형 설정에 대하여 분석 방법을 정리하고, 4장에서는 ERGM을 활용하여 직장 네트워크를 분석한 결과를 정리하며 최종 모형 설정이 적합한지 수렴성을 살펴본다. 마지막으로 5장에서는 결론과 시사점 그리고, 본 연구의 한계에 대해 정리한다.

2. 지수족 랜덤 그래프 모형

2.1. ERGM 특성

네트워크에서 노드(node, vertex, actor)는 각 개체를 의미하며, 에지(edge)는 이러한 노드들의 연결 관계를 나타낸다. 노드와 에지로 구성된 사회연결망 구조를 분석할 때 최근 사용되는 분석으로 ERGM(exponential random graph model) 활용이다 (Kang 등, 2021). 이 방법은 네트워크를 통계적으로 모델링하여 네트워크의 확률 과정(stochastic process)적인 요소를 분석하는 모형이며 네트워크가 몇 가지의 네트워크 원리 및 구성(구조 변수)에 의해서 발생할 수 있는 여러 가지 연결망 중의 하나임을 의미하고, 이러한 방법을 통해 어떠한 원리로 네트워크가 만들어졌는지에 대한 가설검증이 가능하다 (Seo, 2013).

ERGM은 네트워크에서 노드 간 연결을 예측하기 위하여 로지스틱 회귀를 활용한다. 노드 한 쌍의 연결 확률 및 여부(0, 1)가 네트워크 내부의 여러 변수에 따라 예측되어 로지스틱 회귀처럼 로짓(log-odd)으로 풀이할 수 있다. 그러나, 변수 간 독립성을 가정하는 로지스틱 회귀분석 방법과는 달리 실제 사회연결망 분석에서는 노드 간 상호의존성을 고려해야 한다. 예를 들어, 사회연결망의 일반적인 특징인 상호성과 전이성 같은 비독립성은 기존 회귀방식의 가장 기본적인 가정을 위반하며 이를 보완하기 위하여 개인을 사회적 맥락에서 분리하여 독립적으로 취급하는 것은 불가하다. 따라서 본 연구에서는 ERGM을 바탕으로 직장 네트워크를 분석한다.

네트워크의 구조 관계는 ERGM 분석 시에 유의미한 변수로 영향을 끼칠 수 있다. 첫째, ‘상호호혜성(reciprocity)’은 사회적 관계의 가장 기본적인 속성으로, 서로 연결되는 구조이다 (Park, 2019). 본 연구에서는 서로 조언을 구하는 경향으로 해석될 수 있다. 둘째, k -별 구조(k -star)는 한 노드를 중심으로 k 개의 노드가 별 모양으로 전개되는 구조이며, 한 직원이 여러 직원과 조언을 주고받을 관계를 의미한다. 네트워크 데이터의 방향성 정보를 추가하면, 들어오는 방향(in)과 나가는 방향(out) 두 가지로 의미가 확장된다. 그리고 숫자 k 를 지정할 수 있으며, 본 연구에서 ‘조언제공성’은 한 직원이 두 명의 직원에게 조언하는 행위(2-in-star)를

의미한다. 셋째, 이런 지역적인 구조는 확률 모형에서 퇴화(degeneracy) 문제를 일으켜 적절한 모델 적합에 문제가 생길 수 있다 (Kolaczyk와 Csardi, 2014). 이를 위해 삼각형 구조의 퇴화 없이 전이성(transitivity)을 포함할 수 있는 공유 파트너 관계의 기하학적 가중치항(geometrically weighted edgewise shared partnerships; GWESP)이 존재한다 (Hunter 등, 2005). 더불어 직장 네트워크의 조인 제공자와 조인 수용자의 방향성을 고려하기 위해, 조인 삼각형 내 연결의 방향에 따른 다양한 유형의 삼각형을 고려할 수 있는 가중치 항(DGWESP)을 사용한다. 본 연구에서는 A가 B에게 조인을 구하고 B가 C에게 조인을 구하면 A가 C에게 조인할 가능성의 ‘전이성(outgoing two path type; OTP)’을 의미한다. 본 연구의 분석은 프로그램 R의 ERGM 패키지와 함수를 활용한다.

2.2. ERGM 기초모형

본 연구에서 분석할 네트워크 데이터는 방향성을 가지므로, Y_{ij} 는 노드 i 에서 노드 j 로의 연결 여부를 나타내는 변수이며 $Y_{ij} \neq Y_{ji}$ 이다. 본 연구에서 사용한 ERGM 기초모형은 네트워크 구조에 관한 항을 추가하였고, 식 (2.1)과 같다.

$$P(Y = y) = \exp \left\{ \sum_{1 \leq k \leq n-1} \theta_k S_k(y) + \mu M(y) + \tau T(y) + \psi(\theta, \tau) \right\}, \quad y \in Y, \quad (2.1)$$

여기서 $\theta = \{\theta_k, \mu, \tau\}$ 는 모수이고, $\psi(\theta, \tau)$ 는 정규화 상수(normalizing constant)이다. S_k 는 k 값에 따라 연결(edge)과 조인제공성(k -in-star)을 설명하는 개수 통계량으로 식 (2.2), (2.3)과 같다. M 은 상호호혜성(reciprocity)의 구조에 대한 개수 통계량이고, T 는 전이성 구조에 대한 네트워크 통계량으로 식 (2.4), (2.5)와 같다.

$$S_1(y) = \sum_{1 \leq i, j \leq n} y_{ij}, \quad (2.2)$$

$$k\text{-in-star}(k \geq 2) : S_k(y) = \sum_{1 \leq i \leq n} \binom{y_i}{k}. \quad (2.3)$$

$$M(y) = \sum_{1 \leq i < j \leq n} y_{ij} y_{ji}, \quad (2.4)$$

$$T(y) = e^\alpha \left[\sum_{1 \leq a \leq n-2} \{1 - (1 - e^\alpha)^a\} \left\{ \sum_{1 \leq i, j \leq n} y_{ij} I \left(\sum_{1 \leq k \leq n} y_{ik} y_{kj} = a \right) \right\} \right]. \quad (2.5)$$

조인제공성 식 (2.3)에서 y_i 는 조인을 구하는 노드 i 에 조인을 하는 고정된 노드 ‘·’로 연결되는 (i, \cdot) 쌍의 여부를 의미하며 $S_k(y)$ 는 k 개로 조합할 수 있는 경우의 수를 의미한다. 전이성 식 (2.5)에서 양수 α 는 퇴화 모수이며, $\sum_{1 \leq i, j \leq n} y_{ij} I(\sum_{1 \leq k \leq n} y_{ik} y_{kj} = a)$ 는 $i \rightarrow k \rightarrow j$ 관계, 즉 OTP 관계를 맺으며 노드 i 와 노드 j 에 공통으로 연결된 노드 k 의 개수가 a 개인 경우의 수를 의미한다.

ERGM 모형에서는 네트워크 구조 관계뿐만 아니라 노드의 속성에 따라 주효과를 검정할 수 있다 (Hunter 등, 2008). 분석에 사용할 네트워크 데이터가 방향성을 가지고 있다면, 이 정보를 주효과 항에 추가할 수 있으며 노드의 속성이 질적 특성인지 양적 특성을 가지는지에 따라 다르게 설정할 수 있다. 그리고 노드의 특정 속성에 대하여 동질성(homophily) 혹은 이질성(heterophily)의 문제를 생각할 수 있다. 동질성은 서로 유사한 행위자들 간의 연결 성향이 강한 경우이고, 이질성은 서로 다른 행위자들 간의 연결 성향이 강한 경우를 의미한다. 이외에도 네트워크를 구성하는 노드 간의 연결 정도 즉, 더 자주 교류하는 조합일수록 연결 가능성이 커지는 점을 ERGM 모형에 반영할 수 있으며 관계 정보(edgecov)인 가중치 속성으로 연결 정도를 검정할 수 있다.

Table 1: Frequency table of consulting company characteristics

Node	Value	Category	Frequency	Proportion(%)	
Gender	1	남성	Male	34	77.3
	2	여성	Female	10	22.7
Rank	1	연구 조교	Assistant	4	9.1
	2	주니어 컨설턴트	Junior	9	20.5
	3	수석 컨설턴트	Senior	10	22.7
	4	관리 컨설턴트	Managing	17	38.6
	5	파트너	Partner	4	9.1
City	1	보스턴	BOS	24	54.5
	2	런던	LON	1	2.3
	3	파리	PAR	9	20.5
	4	로마	ROM	2	4.5
	5	마드리드	MAD	2	4.5
	6	오슬로	OSL	3	6.8
	7	코펜하겐	CPH	3	6.8
Region	1	유럽	Europe	20	45.5
	2	미국	USA	24	54.5

Table 2: Weight and frequency of ties according to getting advice

Getting advice for 3 months	Weight	Frequency	Proportion(%)	
드물게	Seldom	2	211	40.3
가끔	Sometimes	3	142	27.2
자주	Often	4	84	16.1
매우 자주	Very Often	5	86	16.4

3. 직장 네트워크 데이터

3.1. 연구대상

본 연구에서는 한 컨설팅 회사(consulting company) 내 직원 46명의 네트워크 데이터(Cross와 Parker, 2004)를 활용한다. 연결이 전혀 없는 두 직원을 제외하여 결과적으로 44명의 데이터를 직장 네트워크 분석에 사용하고, 직원들의 기본 정보는 Table 1과 같다. 모든 직원의 속성들은 결측값이 없으며, 범주형 변수의 성별, 도시, 지역 속성과 연속형 변수의 직급 속성을 가지고 있다. 여자 직원보다 남자 직원이 많으며, 직급은 관리 컨설턴트, 수석 컨설턴트, 주니어 컨설턴트, 연구 조교와 파트너 순으로 많다. 각 직급은 Table 1에 적힌 순서대로 연구조교가 가장 낮은 직급, 파트너가 가장 높은 직급이다. 직급 변수는 숫자가 커질수록 높은 직급에 위치한 순서형 변수이므로 연속형으로 처리하여 분석하였다. 지역 변수는 미국과 유럽이 있으며, 미국에 위치한 보스턴 도시에 근무 중인 직원이 가장 많은 것을 알 수 있다. Table 2는 A가 B에게 지난 3개월 간 조언을 구한 정도에 대한 질문에 대해 드물게, 가끔, 자주, 매우 자주로 응답한 경우를 나타내었으며, 정도에 따라 가중치 속성을 부여한 것을 나타낸다.

Figure 1부터 Figure 5까지는 직장 네트워크 분포와 속성을 시각적으로 살펴보기 위한 프루터만-레인골드(Fruchterman-Reingold) 그래프이다. 프루터만-레인골드 알고리즘은 노드끼리는 서로 밀어내는 척력(repulsive force)이 작용하고, 예지가 있으면 서로 당기는 인력(attractive force)이 작용한다는 가정을 바탕으로 노드를 배치하는 방식이다(Fruchterman과 Reingold, 1991). 아래 개별 속성을 표현한 모든 그래프는 가중치 속성

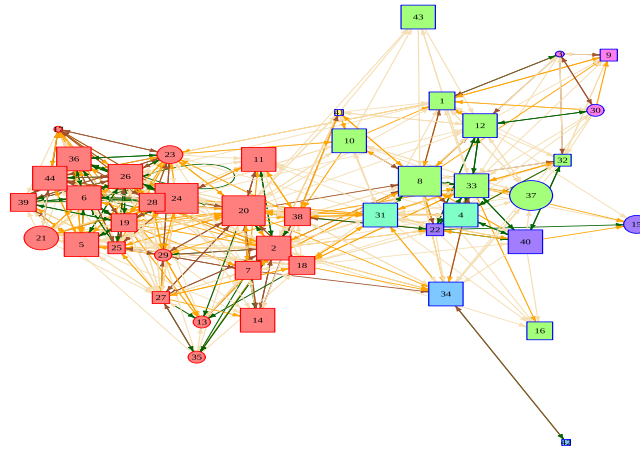


Figure 1: Company network graph including all node and link characteristics.

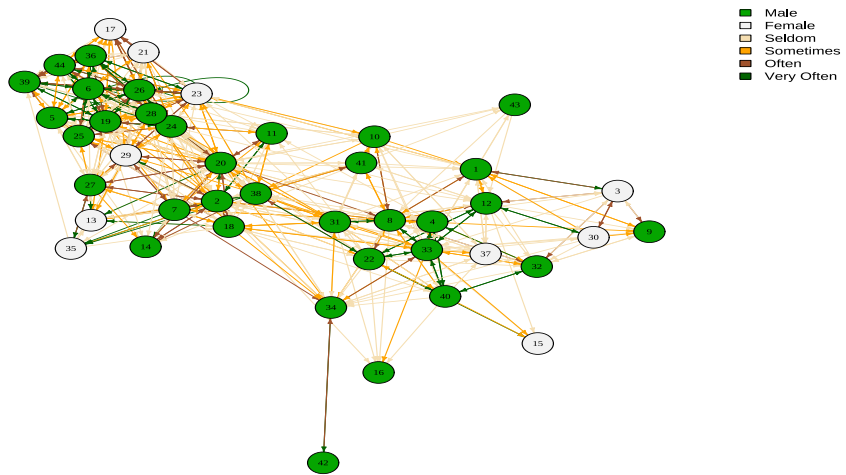


Figure 2: Company network graph focusing on gender characteristics.

정도에 따라 연결 색을 설정하였다. 조언을 많이 얻은 순으로 매우 자주(녹색), 자주(갈색), 가끔(주황색), 드물게(밀색)를 표현하였다. 즉, 녹색과 갈색 연결의 경우 해당 노드 간의 조언 관계가 강하게 형성하였다고 볼 수 있다.

Figure 1은 모든 속성을 표현한 그래프이다. 성별 속성을 구분하기 위해 남자 직원(사각형)과 여자 직원(원)을 다르게 표현하고, 연속형 변수인 직급에 따라 노드 크기를 달리하여 직급이 높을수록 노드 크기를 크게 설정하였다. 지역에 따라 노드 테두리 색을 미국(빨간색)과 유럽(파란색)을 다르게 설정하였다. 도시는 노드의 색으로 표현하였으며 보스턴(빨간색), 런던(노란색), 파리(초록색), 로마(민트색), 마드리드(파란색), 오슬로(보라색), 코펜하겐(분홍색)이다. 예를 들어, 직장 데이터에서 가장 높은 비율을 차지하고 있는 미국 보스턴 지사에 다니는 직원은 빨간색 테두리의 빨간색 노드로 나타난다. 미국의 보스턴 지사에 다니고 있는 남자인 6번, 26번 그리고 39번 직원이 높은 가중치의 화살표와 많이 접하고 있음을 확인할 수 있다. 미국 보

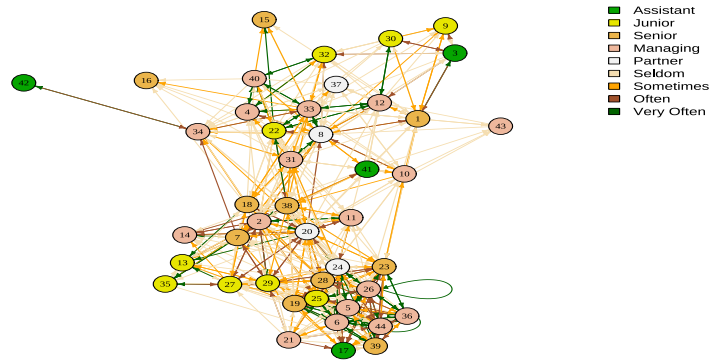


Figure 3: Company network graph focused on rank characteristics.

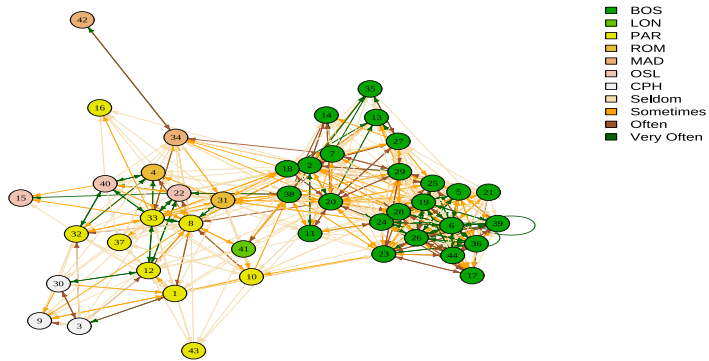


Figure 4: Company network graph focusing on city characteristics.

스턴 지사에 다니고 있는 직원끼리 밀집되어 있음을 볼 수 있는데, 이를 통해 지역 속성의 영향을 짐작할 수 있다. 그 외 남자 직원끼리 연결되어 있는 경우도 다른 성별의 조합보다 많이 보인다.

Figure 2는 성별 속성에 따라 직장 네트워크를 탐색하기 위한 그래프이다. 남자 직원(녹색)과 여자 직원(흰색)을 표현하였다. Figure 2의 그래프에서 흰색보다 녹색 노드가 많으므로, 남자 직원의 수가 여자 직원의 수보다 많음을 확인할 수 있고 여자보다 남자 직원들끼리 밀집되어 있음을 볼 수 있는데 특히 6번, 26번 그리고 39번 남자 직원이 높은 가중치의 연결을 주로 많이 받으면서도 조언을 많이 한 것을 알 수 있다.

Figure 3는 직급 속성에 따라 직장 네트워크를 탐색하기 위한 그래프이다. 가장 높은 직급인 파트너(흰색)부터, 관리 컨설턴트(주황색), 수석 컨설턴트(연한 주황색), 주니어 컨설턴트(노란색), 연구조교(녹색) 순으로 표현하였다. 두 번째로 직급이 높은 관리 컨설턴트에게 조언을 구한 경우가 많고, 앞서 조언을 많이 한 6번과 26번 남자직원 또한 관리 컨설턴트 직원이다.

Figure 4에서 직원의 도시 속성은 보스턴(녹색), 런던(연두색), 파리(노란색), 로마(주황색), 마드리드(분홍색), 오슬로(연한 분홍색), 코펜하겐(흰색)으로 표현하였다. 그리고, Figure 5에서 지역 변수에 따라 미국(녹색), 유럽(흰색)으로 표현하였다. Figure 4와 Figure 5을 통하여 미국 내 보스턴에 근무 중인 직원이 많으며, 그들끼리 가장 많이 교류하고 있다는 것을 알 수 있다. 특히 성별과 직급 변수에 따라 조언을 많이 한 6번과 26번 직원 또한 미국 보스턴 지사의 직원이다.

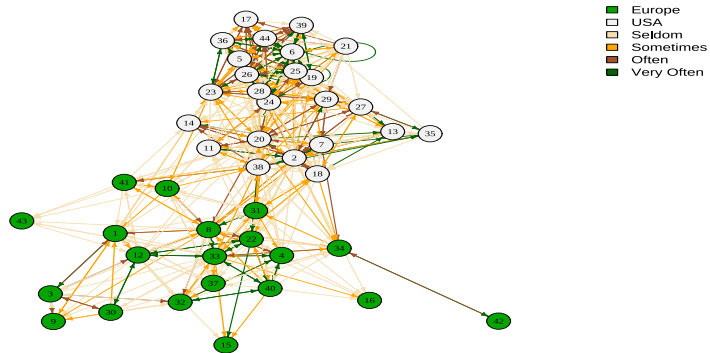


Figure 5: Company network graph focused on region characteristics.

Table 3: Network basic statistics

Clustering Coefficient		0.62
Mean Distance		1.96
Maximum Distance		13
In-Degree	Mean	11.886
	Variance	27.126
Out-Degree	Mean	11.886
	Variance	46.894
In-Strength	Mean	36.682
	Variance	314.455
Out-Strength	Mean	36.682
	Variance	506.362

Table 3은 네트워크의 기초통계량 중 결집 계수(clustering coefficient)(방향성 및 가중성 미반영)와 평균 거리 및 최대 거리를 정리한 결과이다. 네트워크에서 결집 계수란 한 노드가 인접한 노드들과 연결되어 삼중(triangle)을 형성할 확률을 측정하는 통계량이다. 결집 계수는 0과 1사이의 값을 가질 수 있으며 1에 가까울 경우 노드들의 연결이 밀집되어 있다고 해석할 수 있다. 즉, 결집 계수는 네트워크의 밀집된 정도를 보여준다. 직장 네트워크의 방향성과 가중성이 고려되지 않은 결집 계수는 0.62이다.

네트워크의 거리에 대한 일반적인 개념은 노드 간 가장 짧은 길이로 정의한다 (Kolaczyk와 Csardi, 2014). 따라서 평균 거리는 모든 노드 간 최단 거리의 평균을 의미하고, 최대 거리는 노드 간 거리 중 가장 긴 거리를 의미한다. 연결의 방향성과 가중성을 모두 고려한 평균 거리는 1.96, 최대 거리는 13이다. 최대 거리에 비해 평균 거리가 짧게 측정되었으므로 직장 네트워크에 친밀한 관계와 그렇지 않은 관계가 섞여 있음을 짐작해볼 수 있다. 차수(degree)는 특정 노드를 기준으로 들어오거나 나가는 연결의 수를 의미하며, 강도(strength)는 각 연결에 부여된 가중치를 고려해, 들어오거나 나가는 연결의 가중치를 합한 수치로 정의한다. 예를 들어, 노드 A로부터 가중치가 2인 연결 하나와 가중치가 3인 연결 하나가 들어온다면, 노드 A의 들어오는 방향 차수(in-degree)는 2, 들어오는 방향 강도(in-strength)는 5이다. in-degree는 조언 한 직원의 수를 의미하고, out-degree는 조언 구한 직원의 수를 의미한다. 직장 네트워크의 연결은 방향과 가중치를 포함하고 있으므로 Figure 6에서 가중치를 반영한 차수인 강도 즉, 다른 직원에게 조언을 한 정도와 조언을 구한 정도의 합을 히스토그램으로 나타내었다. 조언을 한 정도의 합보다 조언을 구한 정도의 합의 분산이 크고, 편차가 큰 것을 알 수 있다. 직장 네트워크에서 조언을 구할 경우, 다양한 관계의 직원들에게 조언을 구한다는 것을 알 수 있다.

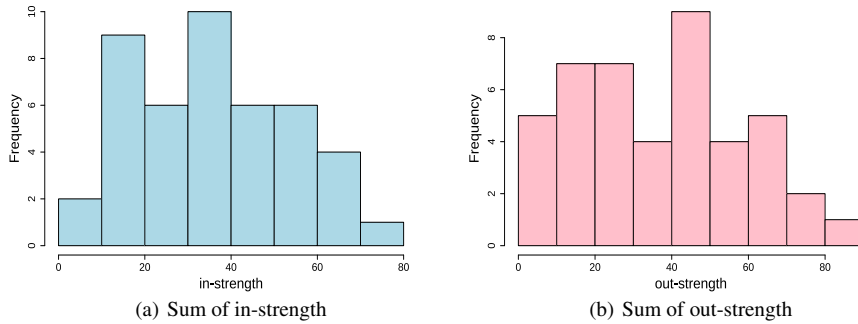


Figure 6: Frequency of sum of strength along the direction.

3.2. 분석 방법

본 연구에서는 ERGM을 활용하여 컨설팅 회사 직원 44명의 네트워크를 살펴보고자 한다. 먼저 직장 네트워크의 구조적 특성을 탐색하기 위해 노드와 에지 간의 관계를 탐색하고자 한다. 해당 네트워크에서 노드는 직원을 의미하고, 에지는 두 직원 간의 연결을 의미한다.

연결은 네트워크 관계에서 가장 기본이므로 먼저 연결만을 고려한 모형을 적합하였다. 다음으로 직장 네트워크에 영향을 주는 요인들을 탐색하기 위해 네트워크의 구조적 특징을 확인하고자 하였다. 따라서 상호호혜성, 전이성, 조건제공성 등 기본적인 구조적 변수를 고려하여 적합하였다. 여기서 상호호혜성이란 직원들끼리 조언을 구하는지를 알아보기 위한 변수이다. 두 직원 A, B 사이에서 A가 B에게 혹은 B가 A에게 조언을 구하는 것보다 서로 조언을 구하는 경향이 있는지 알아보려고 한다. 전이성은 만약 A가 B에게 조언을 구하고 B가 C에게 조언을 구했을 때 A가 C에게 조언을 구할 가능성이 있는지 알아보기 위한 변수이며, 조건제공성 변수에서는 한 직원이 두명의 직원에게 조언을 하는 경향이 있는지 알아볼 수 있다.

이어서 구조적 변수 뿐만 아니라 직원들의 속성 간에 동질성, 연관성 등이 작용하는지 알아보기 위해 관련된 변수를 포함하였다. 직원의 개인적 특성과 네트워크 연결에 대한 가중성을 고려하기 위한 가중치 공변량을 추가하였으며 모든 변수에는 방향성을 구분하였다. 이 중 유의한 결과가 나온 변수들을 바탕으로 모형을 설정하여 분석을 진행하였다. 조언을 구하는 정도가 네트워크 속성과 관계에 영향을 미치거나 직원의 속성에 따라 달라질 것으로 기대하였으나, 확률 예측에 유의한 변수로 선택되지는 않았다. 또한 모형의 적합도는 AIC(Akaike's Information Criterion) (Akaike, 1973)와 BIC(Bayesian Information Criterion) (Schwarz, 1978)를 기준으로, 이 값들이 작을수록 모형의 적합도가 적절하다고 판단하여 최종모형을 선택하였다 (Vega Yon 등, 2021). 이후, ERGM에서 빈번히 일어나는 수렴성 문제를 고려하기 위해 마코프 연쇄 몬테 카를로(Markov chain Monte-Carlo; MCMC) 알고리즘을 기반한 모의실험을 통하여 진단하였다 (Roy, 2020).

4. 연구 결과

4.1. 응답자의 일반적인 특성

가중치를 반영한 차수인 강도는 그 대상 직원 수 즉, 차수에 따라 차이가 있으므로 약한 조언 관계가 여러 번 형성된 것과 강한 조언 관계가 적게 형성된 것의 합이 같게 보일 수 있으므로 잘못 해석하는 것을 방지하기 위하여, 정도의 합의 평균을 구해서 살펴보았다. 직원들 간 조언 관계에 대해 상위 5명을 알아본 결과, 조언을 가장 많이 한 직원은 42, 35, 6, 44, 40번이고, 조언을 가장 많이 구한 직원은 36, 6, 11, 42, 26번 순이다. 이러한 결과를 종합하면 6, 44번 직원이 다수의 직원과 여러 번 조언을 주고 받았는데, 이 두 직원은 미국 보스턴

Table 4: Top 5 nodes according to network statistics

Betweenness Centrality	node 20	node 38	node 8	node 34	node 2
	297.63	166.76	131.99	110.94	97.01
Page Rank	node 20	node 2	node 8	node 24	node 26
	0.01	0.04	0.03	0.03	0.03
Hub	node 26	node 24	node 19	node 20	node 6
	1.00	0.90	0.89	0.87	0.84
Authority	node 6	node 26	node 24	node 19	node 44
	1.00	0.96	0.90	0.84	0.81

지사의 관리 컨설턴트로 높은 직급의 남자 직원임을 알 수 있다. 페이지 랭크(page rank)는 웹페이지의 중요도를 측정해 웹 키워드 검색 결과의 우선순위를 정하는 방법이다 (Kleinberg, 1999). 들어오는 연결의 수와 해당 소스 노드의 중요도를 기반으로 한 아이디어로부터 출발하여 구글 검색에 쓰였던 알고리즘이다 (Brin과 Page, 1998). 본 연구에서는 페이지 랭크가 크게 측정된 노드는 직장 네트워크에서 중요한 역할을 하고 있다고 해석할 수 있다.

HITS(hypertext induced topic selection) 알고리즘을 사용하여 허브와 권위를 구할 수 있다. 본 연구에서는 허브는 상대적으로 많은 연결을 내보내는 노드를 말하며 권위는 많은 연결을 받는 노드를 말한다. 허브와 권위가 높게 측정된 노드가 이 네트워크에서 중요한 역할을 하고 있다고 해석할 수 있다.

Table 4는 직장 네트워크에서 중심성이 높은 직원을 파악하기 위하여 관련 통계량을 구하고, 통계량이 높게 측정된 상위 5명의 직원에 관한 정보이다. 각 통계량의 선정된 상위 직원은 다른 통계량에서도 상위 직원인 경우가 많았다. 특히 20번, 24번, 26번 직원은 4개의 중심성 관련 통계량 중 3개의 통계량에서 상위권을 차지하며 높은 중심성을 보였다. 중요한 시사점은 가장 높은 중심성을 보인 20번, 24번, 26번 직원들의 속성을 파악한 결과, 이 직원들은 비슷한 속성을 가지고 있으며 모두 미국 보스턴 지사에 관리 컨설턴트 혹은 파트너로 높은 직급에 위치한 남자 직원이라는 점이다. 그러므로 직장 네트워크의 조언 관계에 있어 미국 보스턴 지사에 다니는 높은 직급에 위치한 남자 직원이 중요한 역할을 하고 있음을 알 수 있다.

4.2. ERGM 분석 결과

컨설팅 회사 직원 간에 조언을 구함에 있어서 영향을 미치는 변수들을 탐색하기 위해 ERGM의 기초모형 식 (2.1)을 바탕으로 다음의 모형을 적합하였다. 먼저 ERGM의 연결 구조만을 포함해 적합한 식은 Model 1 식 (4.1)과 같다. 연결 뿐만 아니라 네트워크의 구조를 살펴볼 수 있도록 상호호혜성, 전이성, 조언제공성을 고려한 식은 Model 2 식 (4.2)와 같다. 마지막으로, 노드 간 유의한 속성도 고려하여 유의한 변수로 적합한 식은 Model 3 식 (4.3)과 같다.

$$P(Y = y) = \exp\{-0.968S_1(y)\}, \quad y \in Y, \quad (4.1)$$

$$P(Y = y) = \exp\{-10.189S_1(y) + 0.042S_2(y) + 3.258M(y) + 5.751T(y)\}, \quad y \in Y, \quad (4.2)$$

$$P(Y = y) = \exp\{-8.849S_1(y) + 0.041S_2(y) + 2.719M(y) + 3.886T(y) + 1.139a_{ij} + 1.181b_{ij} - 0.400(l_i + l_j) + 0.161(g_i + g_j)\}, \quad y \in Y, \quad (4.3)$$

여기서 Model 3 식 (4.3)의 a_{ij} 는 노드 i 와 노드 j 의 지역 속성이 유럽인가에 관하여 동질성을 파악하는 항이다. 노드 i 와 노드 j 의 지역이 모두 유럽이라면 $a_{ij} = 1$ 이고, 둘 중 하나라도 다르다면 $a_{ij} = 0$ 이다. b_{ij} 는 노드 i 와 노드 j 의 지역 속성이 미국인가에 관하여 동질성을 파악하는 항이다. 노드 i 와 노드 j 의 지역이 모두 유럽이라면 $b_{ij} = 1$ 이고, 둘 중 하나라도 다르다면 $b_{ij} = 0$ 이다. l_i 와 l_j 는 노드 i 와 노드 j 의 직급 속성을 적용하는 항으로

Table 5: Fit results by ERGM models

Variable	Model 1		Model 2		Model 3		
	Estimates	SE	Estimates	SE	Estimates	SE	
Relationship	Edge	-0.968***	0.051	-10.189***	1.334	-8.849***	0.998
	2-In-Star			0.042***	0.010	0.041*	0.017
	Reciprocity			3.258***	0.204	2.719***	0.211
	Transitivity			5.751***	1.112	3.886***	0.852
Homophily	Europe				1.139***	0.139	
	USA				1.816***	0.128	
Covariate	Female				-0.400**	0.049	
	Rank				0.161**	0.151	
GOF	AIC	2229		1694		1403	
	BIC	2235		1717		1448	

* : $p < 0.01$, ** : $p < 0.01$, *** : $p < 0.001$

Table 1 직급의 순서에 따라 1에서 5까지의 값을 갖는다. g_i 와 g_j 는 노드 i 와 노드 j 의 성별 속성을 적용하는 방향으로 남성일 경우 0을 넣고, 여성일 경우 1을 넣는다.

Table 5는 Model 1, Model 2, Model 3을 적합하여 나온 변수에 대한 추정값, 표준오차, p 값에 관한 결과이다. Model 1에서 연결에 대한 가능성은 0.275로 추정되었다. Model 2에서는 모든 구조적 네트워크 변수가 유의하며 연결을 제외한 모든 계수는 양수로 나타난다. 직원 간 상호호혜성이 있을 때, 두 직원에게 조언을 하는 조언제공성이 있을 때, 세 직원 사이의 조언을 구하는 전이성이 있을 때 조언 관계를 형성할 가능성이 높다.

Model 3에서는 구조적 네트워크 변수뿐만 아니라 동질성과 연관성을 살펴볼 수 있는 변수 또한 포함하였고 모두 유의하였다. 직원들의 동질성과 연관성을 탐색한 결과 그 중 지역별 변수가 유의미한 것으로 나타났다. 이에 대한 추정량은 유럽과 미국 둘 다 양수로, 같은 지역의 회사를 다니는 직원들끼리 서로 조언을 구하는 경향이 다른 지역끼리의 관계보다 더 높음을 알 수 있다. 더불어 유럽지사의 직원들보다 미국지사의 직원들끼리의 교류가 활발하다는 것도 알 수 있다. 예를 들어, 파트너 직급의 두 남자 직원이 미국에 근무할 경우 서로 조언을 구할 확률이 0.0044인 반면, 유럽에 근무할 경우 확률은 0.0022로 줄어드는 것을 볼 수 있다. 성별 추정량은 음수로 여자 직원보다 남자 직원이 조언을 주고받는 관계에 관여했을 가능성이 크다. 미국에 근무 중인 파트너 직급의 두 여자 직원이 서로 조언을 주고받을 확률은 0.0020, 같은 조건의 남자 직원과 여자 직원의 확률은 0.0029로 모두 두 남자 직원의 확률인 0.0044보다 낮은 것을 확인할 수 있다. 직급 추정량은 양수로 상위 직급일수록 조언을 주고받을 확률이 높다. 예를 들어, 미국지사에서 수석 컨설턴트로 근무 중인 두 남자 직원끼리 조언을 주고받는 확률은 0.0023이지만, 그 중 한 명이 관리 컨설턴트로 승진했을 때의 확률은 0.0027으로 조언관계를 형성할 확률이 증가하는 모습을 볼 수 있다.

모형적합도의 경우, AIC값과 BIC값이 작은 모형을 선호하므로 Model 1과 Model 2에 비해 Model 3이 적합한 것으로 나타났다. 따라서 Model 3을 최종 모형으로 선정한다.

Figure 7은 최종 모형인 Model 3을 바탕으로 시뮬레이션을 통해 도출한 네트워크 그래프이다. Figure 1과 마찬가지로 성별에 따라 노드 모양(여자:원, 남자:사각형)이 다르고, 직급에 따라 노드 크기를 다르게 설정하였으며 지역에 따른 노드 테두리 색(미국: 빨간색, 유럽: 파란색)을 적용하였다. 다만, 최종 모형의 변수 선택 과정에서 유의하지 않아 포함되지 않은 가중치 속성과 도시 속성은 표현하지 않았다. 직원들의 분포가 미국과 유럽이 확연히 분리되므로, 해당 시뮬레이션 네트워크는 최종 모형에서 가장 크고 유의미한 영향력을 보였던 지역 동질성을 잘 반영하였다. 여자 직원보다 남자 직원이 비교적 네트워크의 중심에 분포하고 있으므로 성별 속성이 잘 반영되었으며, 노드의 크기가 큰 직원들이 중심에 분포하고 있으므로 직급 속성 또한 잘 반영

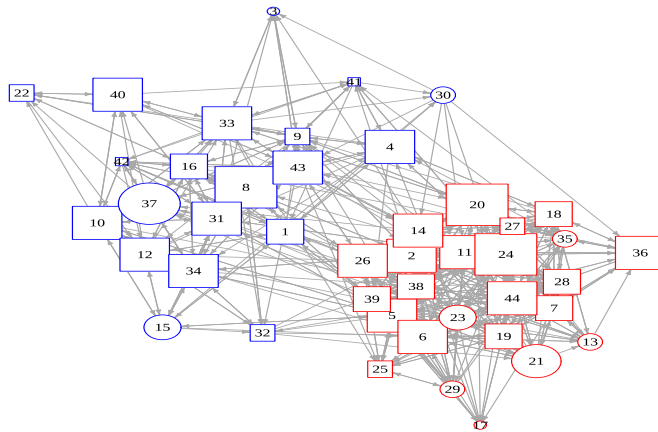


Figure 7: Network simulation graph of the final model (Model 3).

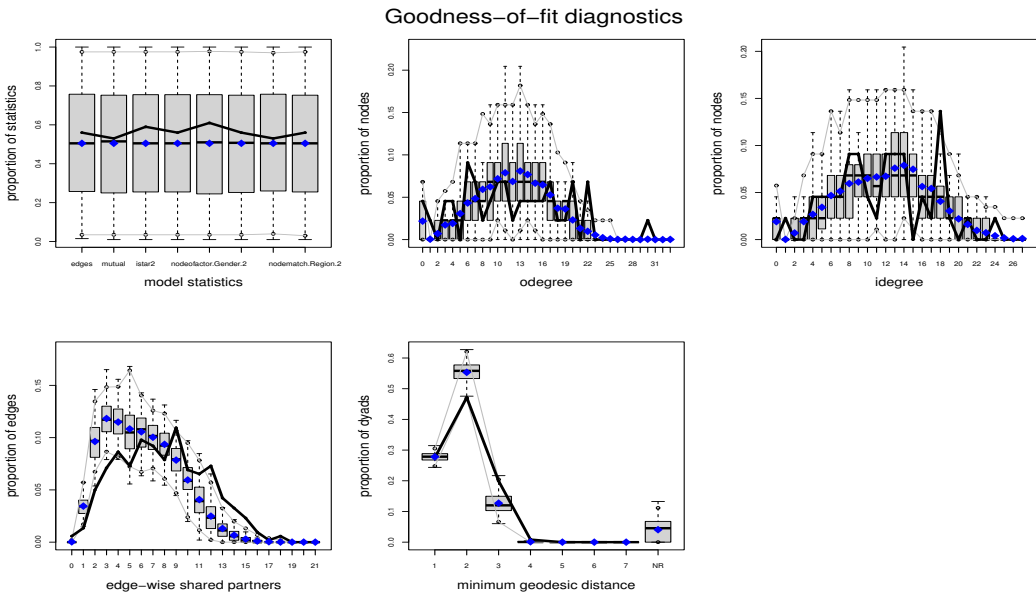


Figure 8: Goodness-of-fit diagnostics for Model 3.

하고 있음을 알 수 있다. 따라서 시뮬레이션 결과를 통하여 최종 모형으로 선택한 Model 3이 직장 네트워크를 설명한 모형이라고 판단할 수 있다.

4.3. 최종모형 진단 결과

마코프 연쇄 몬테칼로(MCMC) 방법은 마코프 체인의 무기역성 성질을 이용하여 확률변수를 시뮬레이션하는 기법이다 (Snijders, 2002). 이 기법을 이용하여 최종모형에서 수렴성 문제가 발생한 경우가 있는지 보았다. Figure 8은 최종 모형인 Model 3의 네트워크 진단 결과이다. 두꺼운 검은색 선은 분석 시 사용한 직장 네트워크

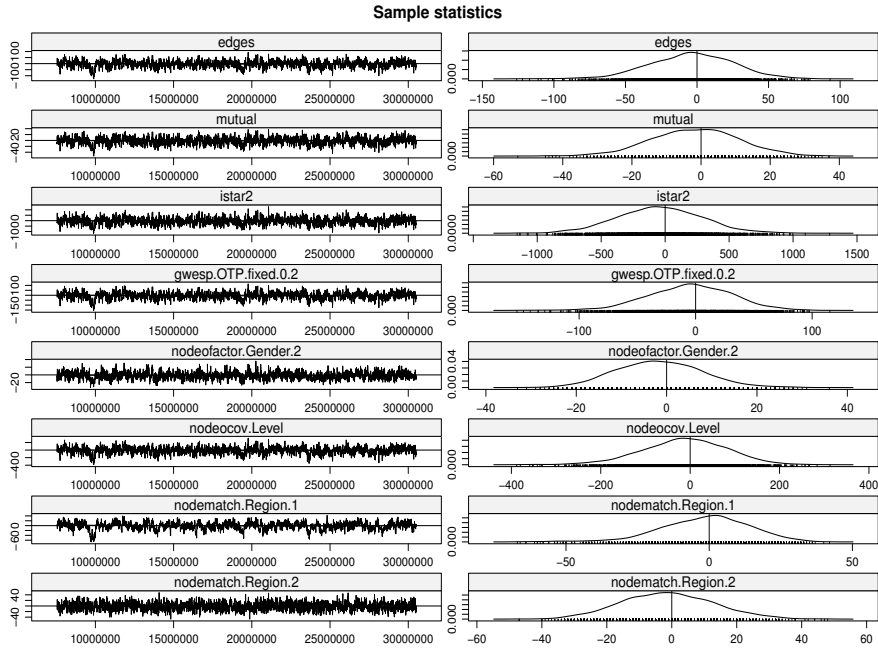


Figure 9: Trends and distributions of differences between MCMC samples and measured values.

의 실측값을 나타내고, 최종모형인 Model 3으로 시뮬레이션을 돌린 결과는 회색 상자그림 분포로 나타나며, 파란색 선은 Model 3의 모의실험에서 얻은 네트워크 통계량의 평균을 보여준다. Figure 8에서 공변량 별 기여도(왼쪽 위), 조인을 구한 노드 개수 비율(가운데 위), 조인을 한 노드 개수 비율(오른쪽 위), 에지 개수 비율(왼쪽 아래) 그리고 양자연결 비율(가운데 아래)을 보여준다. 그림에서 두꺼운 검은색 선이 회색 신뢰구간 안에 있으므로 잘 적합된 모형이라고 볼 수 있다. 또한 Table 5에서 적합도 통계량으로 Model 3의 AIC 값이 세계 모형 중 가장 작다.

Figure 9는 각 공변량에 대해 표본 네트워크들의 통계량 분포를 시간에 따라 보여주고, 이를 통해 최종 모형의 진단 정보를 알 수 있다. 왼쪽 단 그래프는 시간에 따른 표본 통계량의 추이를 알 수 있으며, 오른쪽 단 그래프 그 통계량의 분포를 알 수 있다. 첫 번째 공변량인 연결 공변량에 대하여, 왼쪽 그래프를 보면 실제 데이터와 시뮬레이션 데이터의 연결선 수의 차이가 0 근처에서 변동하고 있으며, 오른쪽 그래프에서 이 차이 들은 0을 중심으로 종모양의 좌우대칭인 정규 분포를 이루고 있다. 다른 공변량들 역시 연결 공변량과 비슷한 결과를 보여주며 이로써 최종 모형이 수렴성을 만족한다는 것을 알 수 있다.

5. 결론 및 논의

본 연구에서는 조인 제공자와 조인 수용자의 속성을 중심으로 직장 내 조인 관계 형성을 파악하고자 컨설팅 회사의 네트워크 데이터를 수집하여 ERGM을 활용하여 분석하였다. 방향성을 고려하여 직원들의 어떠한 속성이 서로 조인을 구하는데 영향력이 있는지 알아보고, 가중성을 고려한 변수, 네트워크 구조에 대한 변수와 노드 속성을 반영한 변수를 ERGM 모형에 적합하여 분석하였다. 특히, 조인 관계에 있어 중요한 역할을 하고 있는 직원을 파악하고, 속성 간에 동질성 혹은 이질성이 존재하는지 알아보았다는 점에서 그 의의가 있다.

본 연구 결과를 바탕으로 제언할 점은 다음과 같다. 첫째, 여러 지사의 구성원보다는 한 지사의 구성원으로

구성할 때, 더욱 잦은 의사소통으로 집단지성의 창의적인 결과를 도출할 수 있을 것이다. 만약 특별한 이유로 여러 지사에 걸쳐 프로젝트 그룹을 구성해야 된다면, 물리적 거리를 좁힐 만한 대책을 마련해 손실을 줄여야 할 것이다. 둘째, 직급이 높은 직원이 의사소통의 주요 구성원일 가능성이 크므로, 낮은 직급끼리의 구성보다 관리 컨설턴트나 파트너 직급의 직원을 포함한 소규모 모임을 구성해 서로 간 효율적인 능력 향상을 도모할 수 있다.

본 연구 결과에서 시사할 점은 다음과 같다. 컨설팅 회사 직장 네트워크는 두 직원 간 조언을 구한 정도를 포함하는 가중 네트워크이다. 하지만 ERGM에 가중치를 포함시켜 분석한 결과, 유의하지 않아 최종모형에서 제외하기로 결정하였다. 따라서 직원들은 서로 조언을 구한 정도가 아닌 단순히 조언을 구한 유무로 해석이 불가피하기 때문에 이 점을 유의할 필요가 있다.

마지막으로, 본 연구의 한계는 다음과 같다. 첫째, 소규모 데이터로 인해 일반화의 오류를 범할 수 있다. 본 연구 결과는 한 특정 컨설팅 회사의 44명 직원의 응답으로 이루어져 있기 때문에 이러한 결과를 일반화하여 다른 회사 직원들에게도 본 연구 결과를 적용한다면 어긋난 결과를 초래할 수도 있다. 둘째, 표본의 편중으로 인한 오류를 범할 수 있다. 본 연구의 데이터는 남자 직원의 비율이 전체의 약 80%이므로 분석 결과에 영향을 미칠 가능성이 높다고 판단된다. 또한, 컨설팅 회사는 미국과 유럽에 지사를 두고 있지만 미국이 속한 도시가 보스턴이 유일한 반면 유럽에는 다양한 도시가 포함되어 있다. 이는 도시의 분포가 지역 간 조언을 주고받음에 영향을 미친다고 할 수 있다. 셋째, 데이터에 포함되지 않은 다른 속성의 영향이 있을 수 있다. 예시로 나이, 학력, 근속연수에 따라 조언을 구하는 정도가 달라질 수 있다. 이 외 변수로 표현하기 힘든 개개인의 성향도 무시할 수 없으며, 이를 고려하는 세밀한 연구도 병행되어야 할 것이다. 넷째, 본 연구에서 분석한 데이터는 특정 시점에서 여러 개체에 대한 조사를 실시하는 횡단면(cross-section) 데이터(Lee 등, 2021)로, 시간의 흐름이 반영되지 않은 정적 데이터이다. Yoon과 Kim (2011)은 횡단조사(cross-section survey)는 한 시점에서 많은 정보를 제공해 줄 수 있으나, 변화를 유발하는 원인의 파악이나 행동 변화의 양태는 보여 주지 못한다는 한계를 가지고 있기 때문에 시간이라는 차원이 추가되어 있어 역동적인 분석이 가능한 패널조사가 필요하다고 하였다. 따라서 동적인 특징을 분석하기 위해선 패널 데이터의 확보와 분석이 요구된다.

References

- Akaike H (1973). *2nd International Symposium of Information Theory*, Akademiai Kiado, Budapest.
- Brin S and Page L (1998). The anatomy of a large-scale hypertextual Web search engine, *Computer Networks and ISDN Systems Journal*, **30**(1), 107–117.
- Brooks AW, Gino F, and Schweitzer ME (2015). Smart People Ask for (My) Advice: Seeking Advice Boosts Perceptions of Competence, *Management Science*, **61**(6), 1421–1435.
- Cross R and Parker A (2004). *The Hidden Power of Social Networks: Understanding How Work Really Gets Done in Organizations*, Harvard Business School Press, Boston.
- Fruchterman TMJ and Reingold EM (1991). Graph Drawing by Force-directed Placement, *Software - Practice and Experience*, **21**(11), 1129–1164.
- Granovetter M (1974). *Getting a Job: A Study of Contacts and Careers*, The University of Chicago Press, Chicago.
- Guk H (2012). A Study on the Personal Tendency and Work Characteristics that Influence Advice Behavior: Focusing on the advice provider's point of view, Korea University, Seoul.
- Hunter DR, Goodreau SM, and Handcock MS (2005). Goodness of Fit of Social Network Models, *CSSS Working Paper*, **47**.
- Hunter DR, Handcock MS, Butts CT, Goodreau SM, and Morris M (2008). *ergm*: A package to fit, simulate and

- diagnose exponential-family models for networks, *Journal of Statistical Software*, **24**(3), 1–29.
- Kang M and Park J (2018). A study on gender differences in organizational network structures, *Gender and Culture*, **11**, 89–123.
- Kang YK, Bae SY, and Hong SH (2021). Analysis of middle school students' friends network in class using ERGM: homophily and relationship in gender, grade, academic achievement and family economic status, *Forum For Youth Culture*, **67**, 5–27.
- Kleinberg JM (1999). Authoritative sources in a hyperlinked environment, *Journal of the ACM*, **46**(5), 604–632.
- Kolaczyk ED and Csardi G (2014). *Statistical Analysis of Network Data with R*, Springer, New York.
- Lee JW, Kang HJ, Oh SH, Cho GW, Jang PS, Seo HJ, Son SA, Na DY, and Song CH (2021). A Study on Developing Panel Data of the Korean Innovation Survey, *Science and Technology Policy Institute*.
- McPherson M, Smith-Lovin L, and Cook JM (2001). Birds of a feather: Homophily in social networks, *Annual Review of Sociology*, **27**(1), 415–444.
- Park HH (2019). Using ERGM in exploring network effects: A case study of policy networks, *Modern Society and Public Administration*, **29**, 35–61.
- Rapoport A and Horvath WJ (1961). A study of a large sociogram, *Behavioral Science Journal*, **6**(4), 279–291.
- Roy V (2020). Convergence diagnostics for Markov chain Monte Carlo, *Annual Review of Statistics and Its Application*, **7**(1), 387–412.
- Schwarz G (1978). Estimating the dimension of a model, *Annals of Statistics Journal*, **6**(2), 461–464.
- Seo IS (2013). A mechanism of collaborative network structure : Focusing on settlement support program, *Korea Research Institute for Local Administration*, **27**, 75–102.
- Snijders T (2002). Markov chain Monte Carlo estimation of exponential random graph models, *Journal of Social Structure*, **3**.
- Song MS (2017). Communication Makes Good Results in the Enterprise, *HR Monthly Review*, **4**, 141–144.
- Stallings MM (2010). Reaching up: The influence of gender, status, and relationship type on men's and women's network preferences, *Publicly Accessible Penn Dissertations*, **144**.
- Vega Yon GG, Slaughter A, and Haye KDL (2021). Exponential random graph models for little networks, *Social Networks*, **64**, 225–238.
- Woehler ML (2017). Gender and Networking: Building and Benignifying from High Status Ties in the Workplace, *Theses and Dissertations–Management*, **8**.
- Yoon YH and Kim HK (2011). Mining and manufacturing panel analysis: Focusing on analysis of the mining and manufacturing industry survey from 2000 to 2009, *Statistical Research Institute*.

Received April 9, 2022; Revised May 9, 2022; Accepted May 16, 2022

직장 네트워크 데이터에 대한 통계적 ERGM 분석

박예진^a, 엄정민^a, 홍수빈^a, 한유진^a, 김재희^{1,a}

^a덕성여자대학교 정보통계학과

요약

회사는 영리 등의 공동 목표를 달성하는 조직으로, 더 나은 성과를 도출해내기 위해 함께 노력하는 수많은 개인으로 구성된 사회 집단이다. 이에 따라 개인의 의사소통 능력을 비롯한 구성원 간의 네트워크 형성이 중요해지고 있다. 이러한 배경으로부터 본 연구는 직원 간 조언 관계 형성에 어떠한 요인이 영향을 미치는지 알아보고자 수행되었다. 이를 위해 미국과 유럽에 지사를 둔 컨설팅 회사 내 직원 44명의 네트워크 데이터를 ERGM(Exponential Random Graph Model) 방법으로 분석하였다. 분석 결과로 첫째, 연결을 비롯해 네트워크의 구조와 관련한 변수들이 유의하였다. 둘째, 서로 조언을 구할 확률에 성별 속성이 가장 큰 주효과로 나타났다. 셋째, 지역별 동질성은 성별 주효과보다 더 큰 연결 확률을 유도하였다. 이러한 결과로부터 직장 내 네트워크가 조금 더 효율적으로 활발하게 이루어질 수 있는 방법을 제시하였다.

주요용어: 네트워크 데이터, 직장 네트워크, ERGM, 상호호혜성, 전이성

이 논문은 정부(과학기술정보통신부)의 재원으로 한국연구재단의 연구 기초연구실 (No. 2021R1A4A5028907) 지원과 기본연구 (No. 2021R1F1A1054968) 지원을 받아 수행한 연구 과제입니다.

¹교신저자: (01369) 서울시 도봉구 삼양로 144길 33, 덕성여자대학교 정보통계학과. E-mail: jaehee@duksung.ac.kr