

Network analysis of urban-to-rural migration

Hyunsoo Lee^a · Jaesun Roh^b · Jin Hwa Jung^b · Woncheol Jang^{a,1}

^aDepartment of Statistics, Seoul National University;

^bDepartment of Agricultural Economics and Rural Development,
Research Institute of Agriculture and Life Sciences, Seoul National University

(Received February 15, 2016; Revised March 25, 2016; Accepted March 30, 2016)

Abstract

Urban-to-rural migration for farming has recently emerged as a new way to vitalize rural economies in a fast-aging rural Korea. In this paper, we analyze the 2013 data of returning farmers with statistical network methods. We identify urban to rural migration hubs with centrality measures and find migration trends based on regional clusters with similar features via statistical network models. We also fit a latent distance model to investigate the role of distance in migration.

Keywords: migration, centrality, HITS algorithm, stochastic block model, latent distance model

1. 서론

경제발전과 함께 우리나라의 주요산업은 끊임없이 변화하고 있다. 이 과정에서 국가 기간산업인 농업은 국가적으로 매우 중요한 산업이지만 전체 경제에서 차지하는 비중은 지속적으로 감소하고 있다. 통계청 경제활동별 GDP 자료에 따르면 농림어업의 GDP 비중은 70년대부터 지속적으로 낮아져 1995년 5.36%에서 2014년 2.11%로 낮아졌다 (Kim, 2014). 농가인구 역시 2000년 403만 명에서 2010년 306만 명, 2014년 275만 명으로 지속적으로 감소하고 있으며, 농업 종사자의 절반 정도가 65세 이상 고령자일 정도로 고령화 현상이 심각하다. 농가인구의 감소와 고령화는 농업의 약화로 이어지며 지역발전에도 걸림돌이 된다. Kim (2010)에 따르면 우리나라 농촌지역에서 일어나고 있는 인구감소와 고령화는 농촌사회의 유지와 지역균형발전에 큰 문제가 될 수 있으며, 따라서 농어촌에 관한 정책은 인구감소와 고령화 문제를 해결하는 데 초점을 두어야 한다.

도시민의 귀농은 이런 농촌문제를 해결할 수 있는 방법 중 하나이다. 기존의 고령화된 농어촌 인구에 비해 귀농인구는 상대적으로 젊은 인적자원이기 때문에 귀농인구의 증가는 농촌의 고령화와 인구감소를 완화시키고 농촌에 활력을 불어넣을 수 있는 기반이 된다 (Roh 등, 2013; Park 등, 2014). 또한 귀농

This work was carried out with the support of “Cooperative Research Program for Agriculture Science & Technology Development (Project No. PJ010718)” Rural Development Administration, Republic of Korea., 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, Korea. E-mail: wjang@snu.ac.kr

Table 2.1. Regional migration flow of returning farmers

지역	유입	유출	순유입	총인구	지역	유입	유출	순유입	총인구
서울	0	2,978	-2,978	10,143,645	경기	1,092	3,189	-2,097	12,234,630
인천	102	872	-770	2,879,782	강원	1,242	354	888	1,542,263
부산	24	1,216	-1,192	3,527,635	충북	1,370	268	1,102	1,572,732
대구	56	957	-901	2,501,588	충남	1,867	247	1,620	2,047,631
광주	0	649	-649	1,472,910	경북	2,895	684	2,211	2,699,440
대전	0	683	-683	1,532,811	경남	1,835	828	1,007	3,333,820
울산	110	279	-169	1,156,480	전북	1,314	493	821	1,872,965
세종	67	1	66	122,153	전남	1,860	295	1,565	1,907,172
					제주	264	105	159	593,806

인의 교육수준이 기존 농업인에 비해 상당히 높고 다양한 직업경험을 갖고 있다는 점에서 (Kang, 2006; Jung 등, 2016; Chae, 2013) 귀농인구의 증가는 농업경쟁력을 높이는 데에도 크게 기여할 것으로 기대된다. 이에 따라 정부차원의 귀농귀촌 활성화 지원을 위해 2012년 국무총리실에서 “귀농귀촌 활성화를 위한 범정부 지원기반 강화”를 위한 “귀농어·귀촌 활성화 및 지원에 관한 법률”을 제정하고 용자 및 세제지원, 주택공급 지원 개선, 교육 및 취업 지원 강화, 지자체 지원사업 등을 추진하고 있다. 정부의 귀농귀촌정책과 지자체 차원의 귀농인 지원정책의 효과를 높이기 위해서는 귀농인의 지역별 이동에 대한 분석과 해석이 필요하다. 지역별로 귀농인의 유입·유출 패턴이 다르다는 점에서 지역별 특성을 고려한 맞춤형 지원정책이 필요하기 때문이다 (Jung 등, 2014).

본 논문에서는 2장에서 자세히 소개할 통계청의 귀농귀촌 데이터를 사용해 귀농인구의 이동을 분석하였다. 인구의 이동에 영향을 미치는 요인은 여러 가지가 있지만 그 중 도착지역에서 누릴 수 있는 이익과 이동에 관한 규제, 그리고 이동거리(실제거리, 문화적 유사도 등)가 도착지를 결정하는데 중요하다고 알려져 있다 (Davis 등, 2013). 귀농인구 이동은 국내에서만 발생하기 때문에 이동에 대한 규제의 효과는 배제할 수 있지만 귀농지역에서 누릴 수 있는 이익(귀농정착금, 지원제도 등)과 이동거리는 귀농인구 이동에 중요한 영향력을 가질 것이라 예상된다. 이를 확인하기 위해 우리는 귀농인구 이동 데이터를 네트워크 모형을 사용해 분석하였다. 네트워크 모형의 관점에서 귀농인이 지역 i 에서 지역 j 로 이동한 것은 vertex i 에서 j 로 향하는 연결인 edge y_{ij} 가 관측된 것이며 이 연결은 방향성(direction)과 이동한 귀농인구수 같은 가중값을 가지고 있다. 본 논문에서 네트워크 데이터를 분석하기 위해 사용한 방법은 크게 vertex의 중요도를 점수화하는 방법과 vertex 간 연결 경향을 추론하는 방법이다. 구체적으로 다양한 vertex centrality 점수를 사용하여 귀농인구 이동 네트워크의 거점지역을 찾고 stochastic block model과 latent space model을 사용해 귀농인구 이동을 설명하는 모형을 제안하였다.

2. 데이터

귀농귀촌에 대한 데이터는 2012년부터 통계청에서 각 행정기관의 자료를 제공받아 귀농귀촌인 규모를 파악, 제공하여 귀농귀촌 정책수립에 필요한 기초자료 제공을 위해 귀농귀촌통계를 수집 및 관리하고 있다. 통계청에 따르면 귀농인은 “동지역에서 읍면지역으로 주민등록을 이전하고 농업경영체, 축산업등록명부, 농지원부에 신규 등록한 농업인”으로, 귀촌인은 “도시에서 농어촌으로 이주한 자”로 정의된다 (통계청, 2014). 통계청은 현재 귀농인구, 지역, 작물과 사육가축별 귀농귀촌에 관련된 데이터를 관리하고 있으며 본 논문에서는 그 중 “2013년 재배작물 가구의 시군구별 현황” 데이터를 사용하였다. 이 데이터는 2012년 11월부터 2013년 10월 사이에 농촌지역으로 이주해 작물을 재배한 귀농인의 수를 조사하고 이를 귀농인이 출발한 지역과 도착한 지역에 따라 기록한 것이다. 이 기간 동안 전국 163개 시와

군 지역에서 14,098명이 귀농해 작물을 재배했으며 도별 현황은 Table 2.1과 같다.

3. 통계적 방법론

3.1. 중심도지표

중심도지표는 네트워크에서 vertex의 중요도를 평가하는 점수로서 일반적으로 지표가 높을수록 그 중요도가 높다. 다양한 중심도지표 중에서 우리는 vertex 사이의 연결유무와 연결의 가중값을 고려하는 지표를 사용해 귀농인구 이동에서 지역의 중요도를 평가하고 귀농 거점지역을 찾아보았다.

3.1.1. Betweenness centrality 네트워크에서 path는 연속적으로 이어진 edge의 나열을 의미하며 betweenness centrality는 네트워크에 존재하는 모든 path 중 특정한 vertex를 지나가는 path의 비율들의 합으로 정의된다. 귀농인구 네트워크에서 지역 j 와 k 가 존재하고 지역 j 가 첫 번째 vertex이고 k 가 마지막 vertex인 path들이 여러 개 존재한다고 가정하자. 이런 path들의 개수를 σ_{jk} 이라 정의하고 이 중 지역 i 를 지나는 path의 개수를 $\sigma_{jk}(i)$ 이라고 한다면 지역 i 의 betweenness centrality는 식 (3.1)과 같이 정의할 수 있다. 따라서 이 점수가 높은 지역은 다른 지역들을 이어주는 경향이 크다고 할 수 있으며, 귀농인구 네트워크에서는 다른 지역으로부터 인구유입과 다른 지역으로의 인구유출이 모두 활발한 지역이라 할 수 있다.

$$C_b(i) = \sum_{j \neq k \neq i} \frac{\sigma_{jk}(i)}{\sigma_{jk}}. \quad (3.1)$$

3.1.2. Eigenvector centrality 네트워크에서 어떤 중요한 vertex가 있고 이런 중요한 vertex들과 연결이 많으면 그 vertex 역시 중요한 vertex라고 하자. 귀농인구 네트워크에서 이런 특성을 지닌 중요도는 중요한 지역으로부터 유입이 많거나 중요한 지역으로 유출이 많은 지역에 높은 점수를 주게 되며 이를 식 (3.2)와 같이 표현할 수 있다.

$$C_e(i) \propto \sum_{j \neq i} (y_{ij} + y_{ji}) C_e(j), \quad (3.2)$$

여기서 y_{ij} 는 지역 i 에서 지역 j 로 이동했는지를 가리키는 지표이며 만약 귀농인구수를 고려할 경우 지역 i 에서 지역 j 로 이동한 귀농인구의 수가 된다. 위의 식을 eigenvector centrality 벡터 \mathbf{C}_e 와 연결상태 행렬 \mathbf{Y} 을 사용하여 식 (3.3)으로 표현할 수 있다. 즉 eigenvector centrality는 연결상태 행렬의 첫 번째 eigenvector로 정의할 수 있다.

$$\lambda \mathbf{C}_e = \mathbf{Y} \mathbf{C}_e. \quad (3.3)$$

3.1.3. HITS algorithm Hypertext-Induced Topic Search(HITS) 알고리즘은 전체 world-wide-web에서 하나의 검색어에 대해 이와 연관된 웹페이지들을 vertex로 가지는 네트워크에서 페이지의 순위를 구하는 알고리즘으로, 연결을 많이 주는 페이지와 연결을 많이 받는 페이지의 순위를 hub점수와 authority 점수에 따라 결정한다. 웹페이지의 링크나 귀농인구의 이동 같은 방향성이 있는 연결로 구성된 네트워크에서 hub점수가 높은 vertex는 다른 중요한 vertex로 연결을 많이 주는 vertex이며 authority 점수가 높은 vertex는 다른 중요한 vertex로부터 연결을 많이 받는 vertex가 된다. 따라서 귀농인구 네트워크의 경우 귀농인구 유출이 많은 대도시에서 인기 있는 귀농지역은 authority 점수가 높고 이들 지역으로 유출되는 귀농인의 출발지역은 hub점수가 높을 것이라 예상할 수 있다.

HITS 알고리즘은 이 두 점수를 계산하기 위해 Algorithm 1과 같은 반복연산을 수행한다. 우선 각 단계에서 지역 i 의 authority점수는 지역 i 로 귀농한 귀농인구의 출발지역의 전 단계 hub점수들의 합으로, hub점수는 해당 지역에서 출발한 귀농인구가 도착한 지역들의 전 단계 authority점수의 합으로 정의된다. 각각의 반복연산 마지막 단계에서 두 점수들은 1과 0 사이의 값을 가지도록 표준화되어 다음 단계의 반복연산으로 넘어간다. 여기서 n 은 귀농인구 네트워크의 지역 개수, P 는 반복연산 횟수를 나타낸다. 또한 E 는 edge 집합을 의미하며 (j, i) 는 지역 j 에서 i 로 향하는 edge를 의미한다 (Kleinberg, 1999).

Algorithm 1: HITS algorithm

```

1. initialization:  $hub^0 = (1, \dots, 1) \in \mathbb{R}^n$ ,  $authority^0 = (1, \dots, 1) \in \mathbb{R}^n$ ;
2. for  $p = 1, 2, \dots, P$  do
    2.1 for  $i = 1, 2, \dots, n$  do
        2.1.1  $authority_i^p = \sum_{j:(j,i) \in E} hub_j^{p-1}$ 
        2.1.2  $hub_i^p = \sum_{j:(i,j) \in E} authority_j^{p-1}$ ;
    end
    2.2 normalize  $hub^p$  and  $authority^p$ 
end

```

3.2. Statistical network models

귀농인구 이동 네트워크의 구조를 살펴보면 지역 간 연결이 귀농인구의 이동에 의해 발생하기 때문에 방향성을 가지고 있다는 것을 알 수 있다. 이러한 방향성을 가진 네트워크 구조 중 가장 널리 사용하는 것은 순환이 없는 방향성 네트워크(directed acyclic graph; DAG)이다. 만약 귀농인구 이동 네트워크가 DAG 구조를 가지고 있다면, 예를 들어 대구에서 경주로 이동하는 귀농인과 경주에서 구미로 이동한 귀농인이 존재할 때 구미에서 대구로 이동한 귀농인은 존재할 수 없지만 2013년 귀농인구 이동 네트워크에는 이와 같은 방향성이 있는 순환이 존재한다. 따라서 우리는 귀농인구 이동 네트워크가 어떤 특정한 네트워크 구조를 가지고 있다고 가정하는 대신 지역 간 연결의 발생이 조건부독립이라고 가정한다. 여기서 조건부독립은 지역에 대한 변수들이 주어진 경우 지역 i 에서 j 로 향하는 귀농인구 이동은 다른 지역 간 이동, 예를 들어 지역 k 에서 l 로 향하는 이동이나 지역 i 에서 k 로 향하는 이동과 독립이라는 것을 의미한다. 귀농인구는 전체 인구에 비해 그 수가 미미하고 한 귀농인의 이동이 다른 귀농인의 이동에 미치는 영향력 또한 적다고 할 수 있기 때문에 이러한 조건부독립 가정이 적절하다고 판단된다.

Latent space model은 네트워크의 조건부독립을 가정하는 모형 중 하나로 관측이 가능한 변수와 함께 관측할 수 없는 잠재 변수(latent variable)를 사용한다. 잠재변수는 특정한 latent space model에 따라 지역별 잠재 그룹을 나타내는 변수가 될 수 있고 지역이 위치한 잠재 공간의 좌표와 이 공간에서의 지역 간 거리가 될 수도 있다. 이런 잠재 변수와 다른 설명 변수들이 주어진 경우, latent space model은 지역 간 이동에 대한 확률이 독립이라 가정할 수 있다. 이러한 모형 중에서 stochastic block model은 잠재 그룹을, latent distance model은 잠재 공간과 그 공간에서의 지역 간 거리를 가정한다 (Snijders, 2011).

3.2.1. Stochastic block model 전국의 모든 지역들이 잠재 그룹 중 하나에 속해있고 같은 그룹에 속해있는 지역들은 다른 그룹으로 이동할 확률이 동일하다고 가정하자. 즉 잠재 그룹 G_q 과 $G_{q'}$ 이

존재하고 지역 i, j 가 G_q 에, 지역 k, l 이 $G_{q'}$ 에 소속되어 있는 경우, 지역 i 에서 k 로 향하는 귀농인구 이동상태 m 의 확률 $\Pr(y_{ik} = m)$ 은 지역 j 에서 l 로 향하는 이동상태의 확률 $\Pr(y_{jl} = m)$ 과 같다고 가정하는 것과 동일하다. 여기서 이동상태 y_{ij} 는 이동여부로 정의하여 Bernoulli 분포를 따른다고 가정한다. 또한 지역 i 가 q 번째 잠재 그룹에 속한다면 잠재 변수 Z_{iq} 의 값은 1이며 $\mathbf{Z}_i = (z_{i1}, \dots, z_{iQ})$ 는 multinomial 분포를 따른다고 가정한다.

Latouche 등 (2012)은 각 지역이 소속된 잠재 그룹과 연결확률을 추정하기 위해 식 (3.4)와 같이 정의된 Kullback Leibler distance를 최소화 하는 방법을 제안하였다. 여기서 marginal log-likelihood의 하한을 크게 하는 것은 식 (3.4)를 줄이는 것과 같으며 이를 위해 \mathbf{Z}, α, π 가 독립이라는 추가적인 가정 하에서 variational EM 알고리즘을 사용한다.

$$\text{KL}(q(\cdot)||p(\cdot|\mathbf{Y})) = - \sum_{\mathbf{Z}} \int \int q(\mathbf{Z}, \alpha, \pi) \log \left\{ \frac{p(\mathbf{Z}, \alpha, \pi|\mathbf{Y})}{q(\mathbf{Z}, \alpha, \pi)} \right\} d\alpha d\pi, \quad (3.4)$$

여기서 $q(\mathbf{Z}, \alpha, \pi) = \Pr(\alpha)\Pr(\pi)\prod_{i=1}^N \Pr(\mathbf{Z}_i)$ 이다. 이 과정은 잠재 그룹의 개수가 주어진 경우이며 잠재 그룹의 개수는 integrated classification likelihood(ICL)를 가장 크게 하는 값으로 결정된다. ICL의 정의는 식 (3.5)와 같다 (Daudin 등, 2008).

$$\text{ICL}(Q) = \max_{\alpha, \pi} \log \Pr(\mathbf{Y}, \mathbf{Z}|\alpha, \pi) - \frac{1}{2} \frac{Q(Q+1)}{2} \log \left(\frac{n(n-1)}{2} \right) - \frac{Q-1}{2} \log(n). \quad (3.5)$$

3.2.2. Latent distance model 이번에는 지역들이 d 차원 잠재 공간에 위치한다고 가정하자. Latent distance model은 이 공간에서 어떤 두 지역 사이의 거리가 가까울수록 해당 지역 사이에서 귀농인구 이동이 발생할 확률이나 이동 귀농인구수(edge weight)의 평균이 증가한다고 가정한다. Handcock 등 (2007)에 따르면, 이동 귀농인구수가 Poisson 분포를 가진 확률변수인 경우 p 개의 지역에 대한 설명 변수 x_{ij} 와 각 지역의 잠재 위치 \mathbf{Z}_i 를 사용해 latent distance model을 일반화선형모형의 형태로 나타낼 수 있다.

$$\Pr(Y = y|\beta, x, Z) = \prod_{(i,j) \in \mathcal{Y}} \Pr(Y_{ij} = y_{ij}|\beta, x, Z) = \prod_{(i,j) \in \mathcal{Y}} \text{Poi}(\exp(\eta_{ij})), \quad (3.6)$$

여기서 η_{ij} 는 $\sum_{k=1}^p x_{k,ij} \beta_k - |Z_i - Z_j|$ 이다. 따라서 지역 i 에서 j 로 이동하는 귀농인구수 y_{ij} 의 평균은 $\exp(\eta_{i,j}(\beta, x_{ij}, Z_i, Z_j))$ 이며 지역 간 잠재 거리가 짧을수록 커지는 것을 알 수 있다. 또한 잠재 위치와 지역의 설명변수가 주어진 경우 각 이동은 독립이라 가정한다.

위의 모형에서 지역의 위치 \mathbf{Z}_i 를 추정하기 위해 사전분포와 parameter의 hyperprior를 다음과 같이 설정할 수 있다 (Krivitsky와 Handcock, 2008).

$$\begin{aligned} \beta_k &\sim N(\xi_k, \psi_k^2), & Z_i &\sim \sum_{g=1}^G \lambda_g \text{MVN}_d(\mu_g, \sigma_g^2 I_d), & \mu_g &\sim \text{MVN}_d(0, \omega^2 I_d), \\ \sigma_g^2 &\sim \sigma_0^2 \text{Inv}_{\chi^2_\alpha}, & (\lambda_1, \dots, \lambda_G) &\sim \text{Dir}(\nu_1, \dots, \nu_G), \end{aligned} \quad (3.7)$$

여기에서 MVN은 d 차원 다변량 정규분포이며 평균 μ_g 의 사전분포는 평균이 $(0, \dots, 0) \in R^d$ 인 다변량 정규분포이다. 또한 $\text{Inv}_{\chi^2_\alpha}$ 는 자유도 α 의 inverse χ^2 분포, $\text{Dir}(\nu_1, \dots, \nu_G)$ 는 Dirichlet 분포를 의미한다. 즉, 지역의 위치 Z_i 는 다변량 정규분포의 혼합모형으로 결정되며 혼합모형의 개수 G 는 미리 정의된 잠재 그룹의 개수 Q 와 같다. 식 (3.8)은 사전분포의 hyperparameter로 \mathcal{Y} 은 관측된 네트워크에 존재하

Table 4.1. Regional vertex centrality measures

betweenness			eigenvalue			hub			authority		
지역	단위	점수	지역	단위	점수	지역	단위	점수	지역	단위	점수
인천	광역시	1762.0	서울	광역시	1.00000	서울	광역시	1.00000	홍천	군	1.00000
김포	시	1122.9	대구	광역시	0.44457	부산	광역시	0.30509	양평	군	0.74393
창원	시	1115.6	인천	광역시	0.43952	대구	광역시	0.30353	단양	군	0.73001
구미	시	932.7	부산	광역시	0.38480	인천	광역시	0.28053	부여	군	0.70093
군산	시	850.2	홍천	군	0.30452	대전	광역시	0.23413	상주	시	0.63095
춘천	시	714.2	의성	군	0.23430	수원	시	0.12453	의성	군	0.62361
경주	시	631.2	상주	시	0.23066	광주	광역시	0.11780	영동	군	0.59410
익산	시	597.9	단양	군	0.22662	성남	시	0.09917	옥천	군	0.57721
포항	시	480.6	양평	군	0.21985	부천	시	0.09292	고창	군	0.54739
대구	광역시	452.7	대전	광역시	0.21691	용인	시	0.09287	횡성	군	0.54560
강릉	시	393.3	부여	군	0.21651	창원	시	0.08176	봉화	군	0.53875
남양주	시	320.7	영동	군	0.19171	고양	시	0.06926	가평	군	0.53392
김천	시	246.8	서산	시	0.19018	안산	시	0.06842	서산	시	0.53091
화성	시	234.5	창녕	군	0.18457	울산	광역시	0.06299	예천	군	0.51917
평택	시	223.7	성주	군	0.18240	안양	시	0.06129	창녕	군	0.49187
파주	시	217.1	봉화	군	0.17729	전주	시	0.05715	거창	군	0.48617
김해	시	214.7	예천	군	0.17693	구미	시	0.05093	논산	시	0.48364
천안	시	214.4	횡성	군	0.17480	광명	시	0.03830	성주	군	0.46381
용인	시	192.5	거창	군	0.17344	청주	시	0.03524	고흥	군	0.45913
경산	시	192.0	김천	시	0.17207	시흥	시	0.03359	밀양	시	0.45695

는 귀농인구 이동 현황 리스트, n 은 전체 지역의 수이다.

$$\begin{aligned} \psi_k &\equiv C_\beta / \left(\frac{1}{|\mathcal{Y}|} \sum_{(i,j) \in \mathcal{Y}} x_{kij}^2 \right), & \sigma_0^2 &\equiv C_{\sigma_0^2} \left(\sqrt{\frac{n}{G}} \right)^2, & \omega^2 &\equiv C_{w^2} C_{\sigma_0^2} \left(\sqrt{\frac{n}{G}} \right)^2, \\ \alpha &\equiv C_\alpha \sqrt{\frac{n}{G}}, & \nu &\equiv C_\nu \sqrt{\frac{n}{G}}. \end{aligned} \quad (3.8)$$

다변량 정규분포의 표준편차 σ 은 \sqrt{n} 에 비례하는데 이는 d 차원 공간에서 n 개의 지역이 위치해 있을 때 그 분산된 정도를 반영하기 위한 것이다 (Krivitsky와 Handcock, 2008). R packages latentnet에서는 $C_{\sigma_0^2} = 1/8$, $C_{w^2} = C_\alpha = C_\nu = 1$ 을 기본값으로 하며 β_k 의 분포는 평균이 0, 분산이 9인 정규분포로 가정한다.

앞에서와 마찬가지로 minimal Kullback-Leibler(MKL) 방법을 사용해 모형의 parameter를 추정한다 (Shortreed와 Handcock, 2006). 구체적으로 MCMC sampling을 사용해 모형과 데이터의 참모형과의 Kullback-Leibler divergence를 최소화 하는 추정값을 구한다 (Handcock 등, 2007).

4. 자료분석

4.1. 귀농 거점지역

먼저 귀농 거점지역을 살펴보기 위해 다양한 중심도지표를 계산하였다. Table 4.1과 Figure 4.1은 앞서 소개한 네 가지 중심도지표에 대해 점수가 높은 지역 순으로 각각 나타낸 것이다.

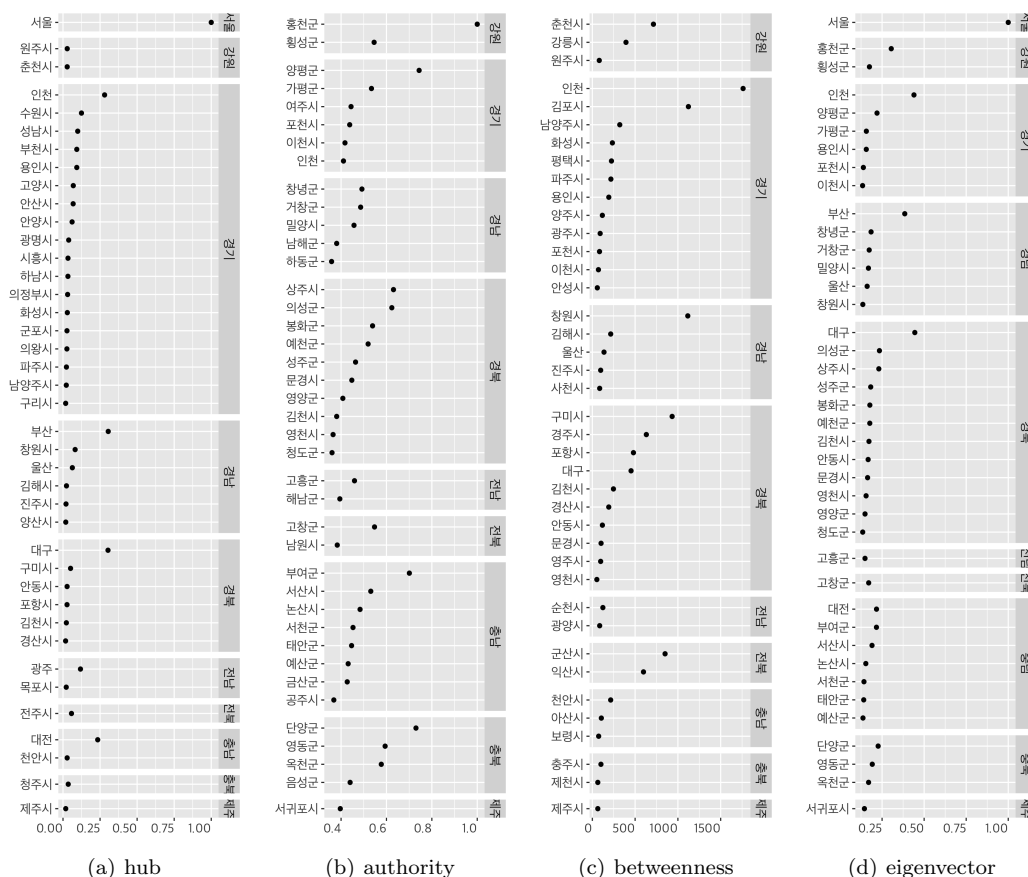


Figure 4.1. Top 40 regions for centrality measures.

먼저 betweenness 점수는 귀농인구의 유입과 유출이 모두 높은 지역인 인천, 김포와 창원 등 대형시에서 높았다. 농촌 지역이 없는 광역시는 귀농인구의 유입이 없고 따라서 path의 중간에 위치할 수 없기 때문에 betweenness 점수는 0이지만, 도농복합도시인 인천, 대구, 울산과 부산은 귀농인구가 유입될 수 있었으며 그 중 인천은 유입인구 103명(서울, 인천, 경기도 대형시 등 9개 지역), 유출인구는 872명(전국 94개 지역)으로 betweenness 점수가 전국에서 가장 높았다. 대구와 울산은 다른 지역에서의 귀농인구 유입이 적게나마 존재하였지만 부산은 지역 내부에서만 이동(같은 지역으로 귀농)이 있었다. 인천 다음으로 betweenness 점수가 높은 지역은 창원, 김포, 구미, 군산, 춘천 등 대체로 각 지방의 중대형 시 지역이었다.

Eigenvector 점수가 가장 높은 지역은 서울, 인천과 같은 광역시이며 다음으로 귀농인구의 유입이 많은 지역은 홍천군, 의성군과 상주시 순이다. 즉 eigenvector 점수는 지역의 유입이나 유출 귀농인구를 함께 고려하는 것을 알 수 있다. 이는 eigenvector 점수와 귀농인구 유출입 합이 상관계수가 높다는 것을 통해서도 확인할 수 있었다.

다음으로 HITS algorithm을 이용하여 hub와 authority 점수를 계산해 보았다. Hub 점수가 가장 높은 지역은 최댓값인 1점을 받은 서울이고 다음으로 부산, 대구, 인천 순이며 귀농인구의 유출이 없는 군

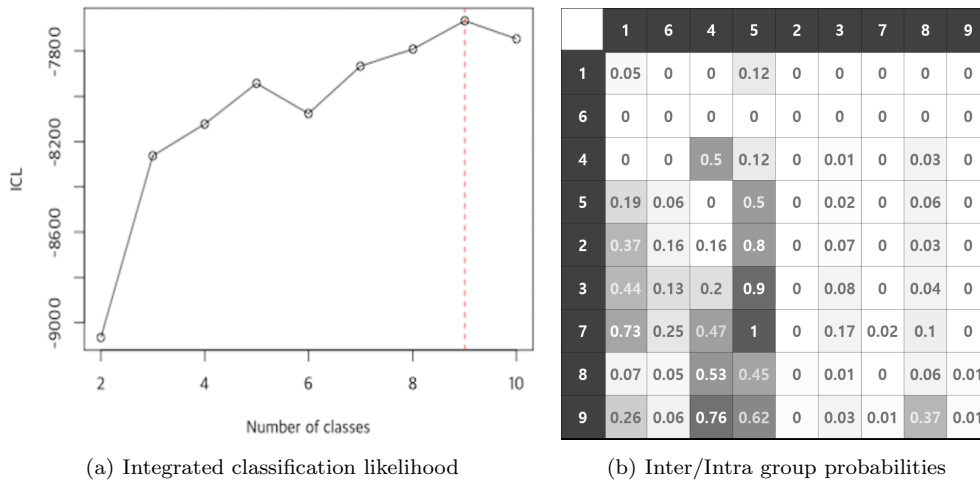


Figure 4.2. Integrated classification likelihood and Inter/Intra group probabilities.

지역은 대부분 0점을 받았다. Hub점수는 각 지역의 유출 귀농인구수와 높은 양의 상관관계를 가지는데 유출인구가 가장 많은 서울을 포함하는 경우 Pearson 상관계수는 0.991이며 포함하지 않는 경우는 0.981이었다. Hub점수에 영향을 미치는 또 다른 요인은 유출 귀농인구가 정착하는 도착지역의 다양성과 도착지역의 authority 점수를 들 수 있다. 예를 들어 hub점수의 순위를 살펴보면 부천시와 창원시는 각각 10위(0.089)와 11위(0.085)이지만 유출 귀농인구수는 부천이 280명, 창원이 404명으로 순위가 바뀌었으며 이들 지역에서 유출된 귀농인의 정착지역은 부천이 60개 지역, 창원이 37개 지역으로, 비록 부천의 전체 유출인구수는 작지만 더 다양한 지역으로 이동하였음을 알 수 있었다. 또한 유출 귀농인구의 도착지역의 authority 점수를 유출인구수에 따라 합할 경우(해당지역 authority 점수와 해당지역으로 유출된 귀농인구수의 곱) 부천이 112.6점, 창원이 99.1점이다. 따라서 도착지역의 authority가 높고 귀농인구의 유출이 많이 발생하는 도시의 hub 점수가 높은 경향을 보임을 알 수 있었다.

Authority 점수는 강원도 홍천군이 최댓값인 1점으로 가장 높으며 다음으로 경기도 양평군, 충북 단양군, 충남 부여군 순이다. Authority 점수는 유입 귀농인구수와 높은 양의 선형상관관계를 보이는데 authority 점수와 유입 귀농인구수의 Pearson 상관계수는 홍천군을 포함하면 0.873이며 제외할 경우 0.875이었다. Authority 점수의 순위를 살펴보면 양평군과 단양군이 각각 2위(0.744)와 3위(0.730)를 차지했지만 총 유입 귀농인구는 각각 12개 지역에서 137명, 23개 지역에서 213명이며, 이 중 서울에서 각각 83명과 67명이 유입되었다. 귀농인구 유입은 단양군이 양평군보다 더 다양한 지역에서 더 많이 발생했지만 hub 점수가 높은 서울에서 인구유입이 많았기 때문에 양평군의 authority 점수가 더 높아진 것으로 보인다. 요약하자면 hub 점수의 경우 다양한 지역으로 연결되는 것이 중요하지만 authority 점수는 핵심적인 지역과 연결되었는지의 여부가 더 중요하다고 할 수 있다.

4.2. 네트워크 모형

4.2.1. 귀농지역 그룹화 Stochastic block model을 사용해 할 경우 지역 간 이동의 존재유무를 이용하여 비슷한 연결성향을 가진 지역들을 그룹화하고 이들의 특징을 파악할 수 있다. 이 모형을 귀농인구 이동 데이터에 적합한 결과 ICL을 기준으로 전체 지역을 9개의 잠재 그룹을 나눌 수 있었다. Figure 4.2(a)는 잠재 그룹 개수에 따른 integrated classification likelihood의 값으로 그룹이 9개일 때 가장 높

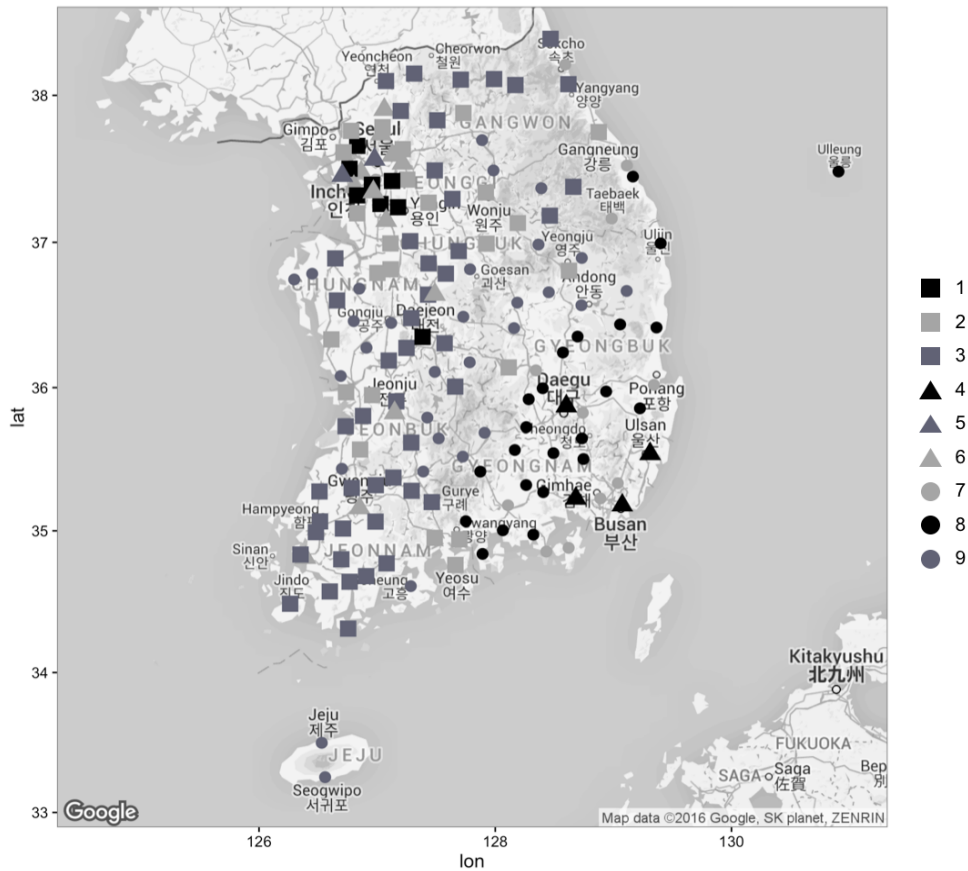


Figure 4.3. Latent group distribution.

다. Figure 4.2(b)는 그룹 간 이동확률을 나타낸 것으로, i 그룹에서 j 그룹으로 이동할 확률은 i 번째 행의 j 번째 열이다. 예를 들어 그룹 5에서 그룹 1로 귀농인구 이동이 발생할 확률은 0.12이다. 그룹 간 귀농인구 이동확률에 따라 타 그룹으로 유출과 그룹 내 이동이 많은 그룹(4, 5번 그룹), 타 그룹에서 유입이 많은 그룹(2, 3, 7, 8, 9번 그룹), 타 그룹으로 유출이 많고 유입은 없는 그룹(1, 6번 그룹)으로 구분할 수 있다.

Figure 4.3은 각 지역이 소속된 잠재 그룹을 표시한 것으로 이를 소속된 지역의 특성과 특정 그룹으로의 이동할 확률에 따라 세 가지 군으로 분류할 수 있다. 첫 번째 그룹군은 부산, 대구, 울산과 창원시로 구성된 4번 그룹과 서울과 인천으로 구성된 5번 그룹이다. 군 단위 행정구역이 없어 귀농인구의 유입이 없는 서울을 제외한 인천, 부산, 대구와 울산은 귀농인구의 유입이 활발한 광역시이다. 이 지역들은 군 단위 행정구역이 존재하며 대부분 내부이동이 많다는 것이 반영되어 4번, 5번 그룹의 그룹 내 이동확률은 0.5로 높은 편이지만 다른 그룹에서 유입이 발생할 확률은 매우 낮아 보인다.

두 번째 그룹군은 1번과 6번 그룹으로 1번 그룹은 수원시 같은 경기도의 대형시 일부와 대전으로 구성되어 있고 6번 그룹은 광주와 전주시, 청주시, 하남시와 나머지 경기도 대형시로 구성되어 있다. 이 지역들은 유입 귀농인구수가 매우 작는데, 46명의 귀농인구가 유입된 용인시를 제외한 다른 지역들은 유

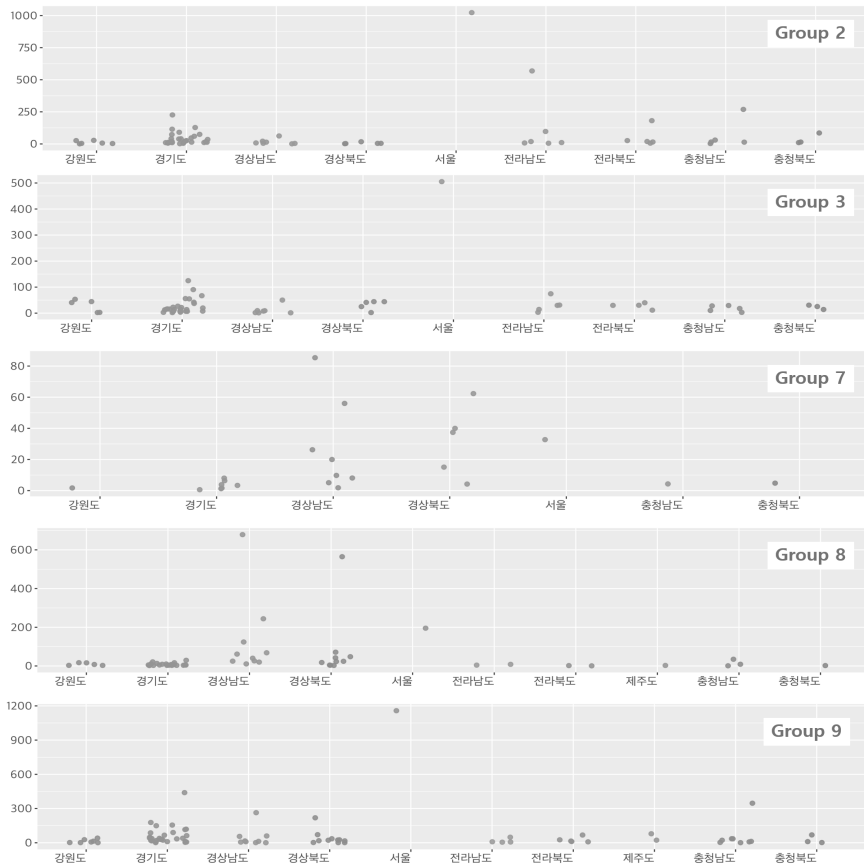


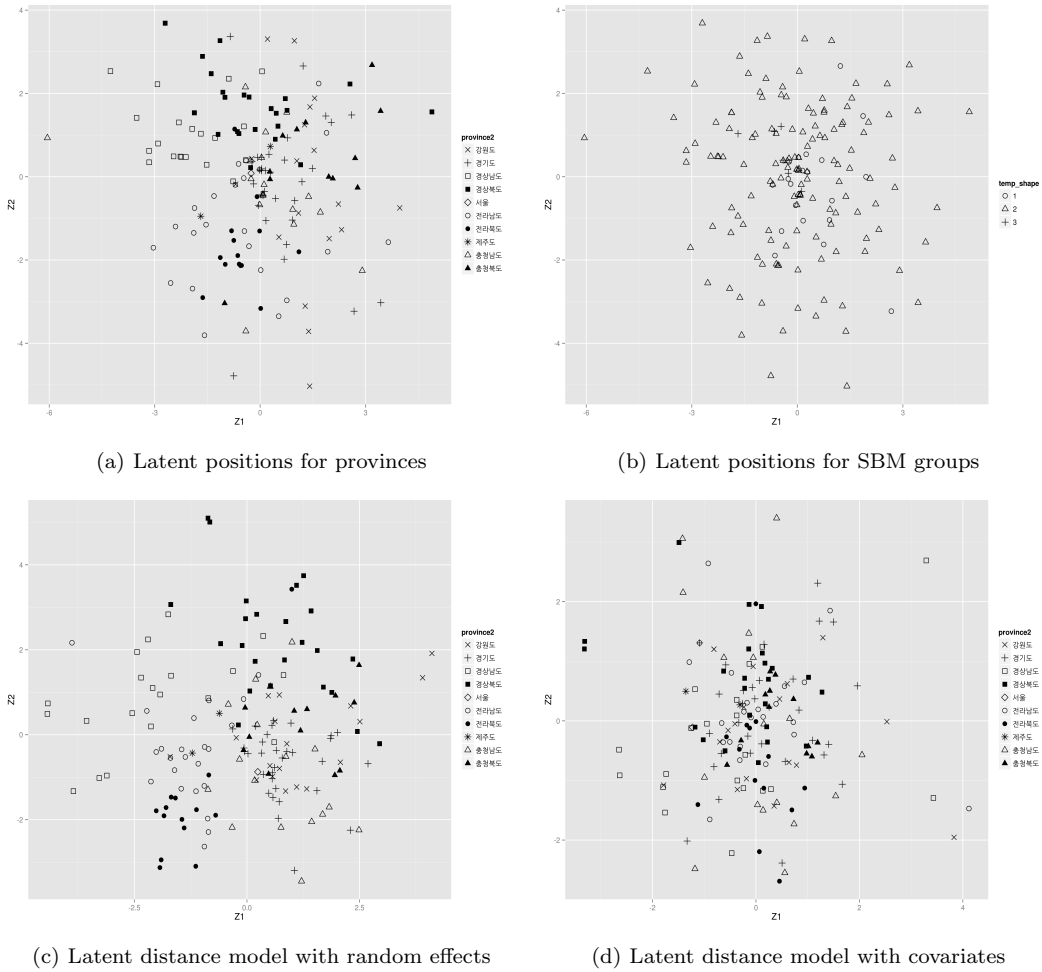
Figure 4.4. Distributions of regions before migration for the 3rd latent group.

입 귀농인구가 존재하지 않는 반면 2번, 3번, 7번 그룹으로 귀농인구의 유출이 활발하였다.

세 번째 그룹군은 시와 군 지역으로만 구성된 나머지 그룹들로서 귀농인구의 유출확률은 매우 낮고 유입 확률이 높은 공통점이 있지만 해당 지역으로 유입되는 귀농인구의 출발지역(그룹)이 차이를 보였다. 예를 들면 5번 그룹(서울과 인천)으로부터 유입확률이 높은 그룹은 2번, 3번, 7번, 9번 그룹이고 4번 그룹(경상도 지방의 광역시)으로부터 유입확률이 높은 그룹은 7번, 8번, 9번 그룹이다.

이를 자세히 확인하기 위해 세 번째 그룹군으로 유입된 귀농인구의 출발지역(유출지역)을 살펴보면 Figure 4.4와 같다. 다른 그룹의 경우 서울에서 유입되는 귀농인구가 가장 많은 반면 7번과 8번 그룹의 경우에는 경상도 지방에서 유입되는 귀농인구가 많았다. 요약하면 2번은 서울과 전라도 지방의 대도시(광주 등)에서 유입이 많은 전국 군, 3번 그룹은 서울에서 유입이 많은 전국 시 지역으로 설명할 수 있으며 7번 그룹은 서울에서 유입이 많은 경상도 지방의 시, 8번 그룹은 경상도 지방의 대도시에서 유입이 많은 경상도 지방, 9번 그룹은 경상도 지방의 대도시에서 유입이 많은 전국의 지역으로 설명할 수 있다.

4.2.2. 귀농인구 이동과 거리 앞선 분석을 통해 귀농인구의 이동이 무작위가 아니라 출발지역과 도착지역의 인접한 정도나 지역 간 거리에 따른 경향이 있음을 확인했다. 지역 간 거리가 귀농인구 이동에



(a) Latent positions for provinces

(b) Latent positions for SBM groups

(c) Latent distance model with random effects

(d) Latent distance model with covariates

Figure 4.5. Latent positions.

미치는 영향을 더 자세히 알아보기 위해 latent distance model을 고려하였다. 귀농인구 네트워크의 경우 edge y_{ij} 는 지역 i 에서 지역 j 로 이동하는 귀농인구의 수이기 때문에 Poisson 분포를 가정할 수 있으며 이동 귀농인구수의 평균은 $\exp(\sum_{k=1}^p x_{k,ij}\beta_k - |Z_i - Z_j|)$ 이라고 가정한다. 잠재 공간에서 지역 간 잠재 거리는 2차원 Euclidean distance를 사용하였다.

만약 추가적인 설명변수 없이 지역 간 잠재 거리만 사용한다면 지역 i 에서 j 로 이동하는 귀농인구수의 평균은 $\exp(-|Z_i - Z_j|)$ 이 된다. Figure 4.5(a)는 설명변수를 포함하지 않는 모형의 잠재 공간에서 지역의 위치로, 같은 도에 소속된 지역끼리 뭉쳐있으며 특히 경상도와 전라도 지방의 지역에서 그 경향이 뚜렷하다. 따라서 latent distance model에서도 귀농인구는 인접 지방으로 이동하는 경향이 크다고 할 수 있다. 서울과 경기지방의 지역은 다른 지역들과의 거리가 가장 가까운 잠재 공간의 중심에 위치해 있으며 이는 이 지역들과 타 지역들 사이에서 귀농인구 이동이 활발히 일어나고 있음을 의미한다.

Figure 4.5(b)는 설명변수를 포함하지 않은 모형에서 지역들의 위치를 stochastic block model에서 분

Table 4.2. Inference results for 4 latent distance models with G groups

Model	Minimal Kullback-Leibler			Var(random effect)	
	절편	실제거리	인구수 차이	sender	receiver
A ($G = 1$)	-2.53501	.	.	19.74261	7.37838
B ($G = 1$)	-1.69358	-6.70997	.	18.30758	7.60880
C ($G = 1$)	0.48524	-2.63866	-26.12422	.	.
D ($G = 1$)	-1.97101	-6.36371	0.41657	13.41403	4.70443
A ($G = 2$)	-2.45469	.	.	18.97966	6.94292
B ($G = 2$)	-1.87547	-6.94694	.	25.78624	7.59060
C ($G = 2$)	1.45537	-2.31819	-28.35297	.	.
D ($G = 2$)	-2.09698	-6.83573	0.90003	20.57696	7.42112
A ($G = 3$)	-2.54357	.	.	22.93055	9.09061
B ($G = 3$)	-2.47230	-6.88926	.	28.52714	7.59670
C ($G = 3$)	1.26952	-3.07253	-36.44428	.	.
D ($G = 3$)	-1.97517	-7.09686	0.43519	28.26001	7.93333
A ($G = 4$)	-2.86601	.	.	27.18794	10.32904
B ($G = 4$)	-2.33291	-6.95139	.	23.29609	7.39786
C ($G = 4$)	1.23368	-2.80396	-35.59539	.	.
D ($G = 4$)	-2.18582	-7.94019	-0.30540	25.57273	7.80906

류한 세 가지 그룹군에 따라 표시한 것이다. 여기서 십자형은 귀농인구 유출과 그룹 내 이동이 많은 그룹, 원형은 유출이 많은 그룹, 삼각형은 외부로부터 유입이 많은 그룹이다. 타 지역으로 유출이 많은 4번, 5번과 1번, 6번 지역은 중심에 위치해있고 나머지 지역은 주변에 흩어져있는데 이것은 유출이 많은 지역을 중심에 두어 타 지역과의 잠재 거리를 가깝게 한 것으로 해석할 수 있다.

이와 같은 설명변수를 포함하지 않은 모형에서 귀농인구수 평균은 두 지역의 잠재 거리에 따라 결정된다. 따라서 지역 i 에서 j 로 향하는 이동과 j 에서 i 로 이동의 평균이 같기 때문에 이동의 방향성을 설명할 수 없다. 이러한 문제점을 해결하기 위해 각 지역별 귀농인구 유출과 유입경향을 나타내는 sender random effect δ_i 와 receiver random effect γ_j 를 모형에 포함시킬 수 있다. 여기서 δ_i 가 큰 지역은 귀농인구의 유출경향이 크고 γ_j 가 큰 지역은 유입경향이 크다. 일반적으로 이들의 분포는 평균이 0인 정규분포를 따르며 서로 독립이라고 가정한다 (Krivitsky 등, 2009). 위의 모형에 두 가지 random effect를 포함할 경우 이동 귀농인구수의 평균은 $\exp(\delta_i + \gamma_j - |Z_i - Z_j|)$ 이 되어 이동 귀농인구수의 평균은 방향에 따라 달라지기 때문에 방향성을 설명할 수 있다.

Figure 4.5(c)는 random effect를 포함하는 모형에서 지역의 위치를 나타내며 방향성을 가지고 있다. 전라도 지방(원형)과 경상도 지방(사각형)은 서로 인접해있으며 경기도 지방(십자형)과 강원도 지방(x형)은 서로 섞여있다. 설명변수를 포함하지 않은 모형 (Figure 4.5(a))과 비슷하지만 전체적으로 더 넓게 퍼져있으며 같은 지방에 속한 지역끼리 뭉쳐있는 정도도 크다. 또한 잠재 공간에서 지역들이 소속된 각 지방의 위치를 살펴보면 실제 도별 위치와 비슷하게 경남이 전남과 경북에 인접해있고, 경북 지방이 경남과 강원, 충북과 인접해 있다.

마지막으로 random effect 모형에 지역 간 실제거리와 인구수 차이를 설명변수로 추가하는 것을 고려해 보았다. 변수 x_{ij} 는 두 지역 i, j 사이의 실제거리를 표준화한 것으로, 이 변수를 포함할 경우 이동 귀농인구수의 평균은 $\exp(\beta_1 x_{1,ij} + \beta_2 x_{2,ij} + \delta_i + \gamma_j - |Z_i - Z_j|)$ 와 같다. Figure 4.5(d)는 이 모형에서 지역의 위치를 나타내며, 대부분 지역들이 정규분포의 중심부에 모여 있다. Figure 4.5(c)와 비교해 특정 패턴이 존재하지 않고 무작위로 섞여있으며 대부분 지역이 정규분포의 중심부에 몰려있어 지역 간 잠재

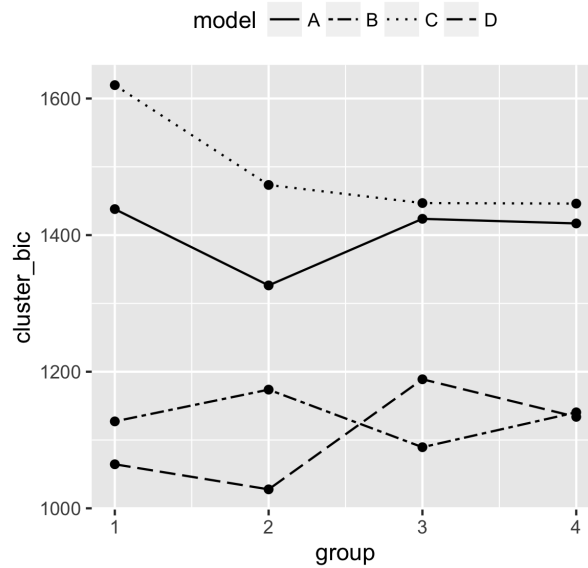


Figure 4.6. BIC for each model.

거리가 가깝다. 따라서 지역 간 실제거리와 random effect가 설명변수로 추가되면 잠재 거리의 영향력이 감소한다고 볼 수 있다.

각 설명변수의 포함여부에 따라 Model A(null model), Model B(실제거리, random effects), Model C(실제거리, 인구수 차이), Model D(실제거리, 인구수 차이, random effects)와 같은 4가지 모형을 고려하였으며 Table 4.2에서는 그룹의 개수별로 각 모형별 추론 결과를, Figure 4.6에서는 각 잠재 그룹 개수별 각 모형의 BIC값을 제시하였다. 동일한 모형을 사용하였을 경우 BIC를 사용해 잠재 그룹의 수를 정할 수 있으며, 분석에서는 그룹의 수가 1-4개인 모형들 BIC값을 비교하였다. 그 결과 Model A와 D의 경우에는 그룹의 수가 2개, B는 4개, C는 3개일 때 가장 작은 BIC를 가졌다.

Model C의 경우 그룹 개수 G 에 따른 추정량의 변화는 크지 않았으며 지역 간 실제거리의 계수 β_1 의 추정값은 음수였다. 따라서 다른 조건이 동일할 경우 두 지역 간 실제거리가 멀수록 평균 귀농인구수가 $\exp(\beta_1)$ 만큼 줄어든다고 해석할 수 있다. 인구수 차이의 계수 β_2 의 추정값 또한 음수인데 인구수 차이 변수를 (유입지역 인구수 - 유출지역 인구수)로 정의하였기 때문에 인구수가 많은 지역에서 인구수가 적은 지역으로 이동하는 경향이 존재한다고 할 수 있다. 실제로 대부분의 귀농인구는 광역시나 시 지역에서 군 지역으로 이동하기 때문에 이 추정값은 타당하다고 할 수 있다.

Model D에서 주목할 점은 지역별 random effect와 인구수 차이에 대한 계수 추정값이 영향을 미친다는 것이다. 지역별 random effect가 포함될 경우 인구수 차이에 대한 계수 추정값의 평균은 -30에서 0에 가깝게 높아지며 인구수 차이가 포함되면 random effect 역시 변화함을 알 수 있었다. 지역의 random effect는 해당 모형에 포함된 설명변수로는 설명할 수 없는 edge의 방향을 설명하는 변수인데, 인구수 차이가 포함되면 $G = 4$ 일 경우를 제외하고 random effect의 분산이 작아지기 때문에 인구수 차이가 귀농인구 이동의 방향성을 설명하는 변수로 추측할 수 있다. 단 random effect의 분산이 변화하는 정도는 그룹의 개수에 따라서도 그 정도에 차이가 있었다. 따라서 귀농인구 이동의 방향과 평균은 지역 간 거리와 인구수 차이 이외에 다른 요인에 의해서 영향을 받는다고 추측할 수 있다.

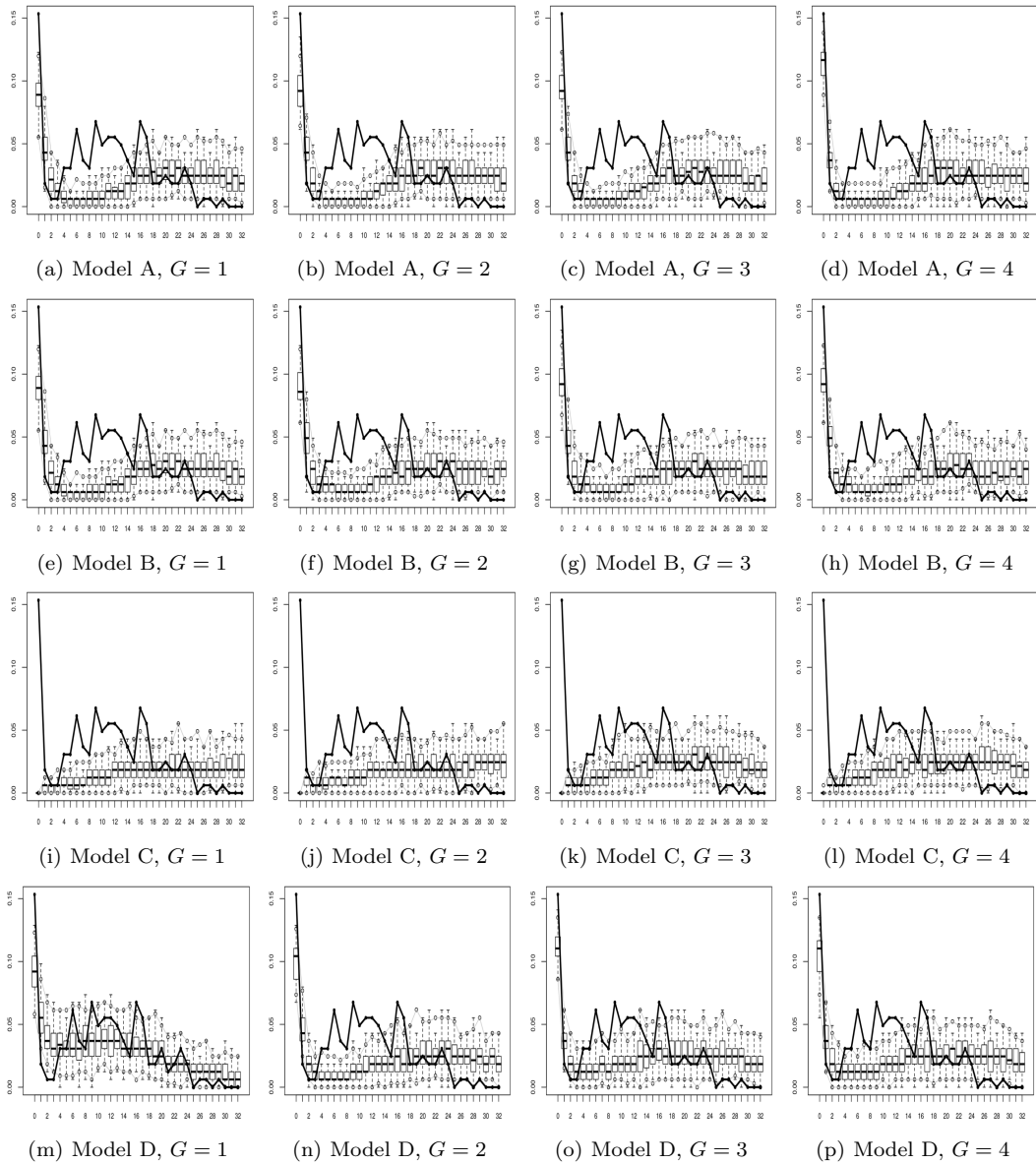


Figure 4.7. Goodness-of-fit plots for in-degree distributions.

마지막으로, 모형의 goodness-of-fit을 통해 여러가지 모형의 적합도를 평가해보았다. Latent distance model에서는 적합된 모형을 사용해 여러가지 네트워크 통계량의 사후 예측분포와 실제 네트워크의 분포를 비교하는 방법을 사용해 적합도를 평가할 수 있으며, 이 때 사용할 수 있는 통계량에는 in-degree, out-degree와 minimal geodesic distance 등이 있다. 이 중 out-degree와 minimal geodesic distance 예측분포는 모든 모형에서 비슷했기 때문에 in-degree의 분포를 통해 적합도를 비교하고 Figure 4.7에 그 결과를 제시하였다. 여기서 굵은 실선은 관측된 in-degree distribution을 나타내며 각 모형의 posterior

distribution에 생성된 샘플들을 이용하여 degree별 boxplot들을 생성하였다.

잠재 그룹의 수가 1인 Model D를 제외한 나머지 모형은 비슷한 in-degree 예측 분포를 보여줬지만 실제 귀농인구 이동 네트워크와는 다른 점이 많았다. 이 모형들은 in-degree의 차수가 높아질수록 그 빈도가 늘어날 것이라 예측했지만 실제 네트워크에서는 차수가 18 이상이 되면 그 빈도가 점점 줄어들며 in-degree의 차수가 23 이상인 지역은 빈도수가 매우 낮다. 이런 경향은 잠재 그룹의 수가 1인 Model D가 비교적 잘 예측했으며, 이것은 귀농인구 이동의 그룹화를 Model D에서 사용한 두 가지 설명변수와 지역의 random effect를 사용해 어느정도 설명할 수 있다는 근거가 된다. 또한 stochastic block model의 경우 이동유무만 고려하지만 latent distance model의 경우 실제 이동인구를 edge weight로 고려하여 그룹의 개수가 다르게 나온 것으로 생각되며 Model D의 BIC값을 살펴보면 그룹의 개수가 1일 경우와 2일 경우 큰 차이가 나지 않으므로 그룹의 개수를 1로 보는 것이 큰 무리가 없다고 생각할 수 있다.

5. 결론

본 논문에서는 전국의 귀농인구 이동을 네트워크 모형을 사용해 분석해보았다. 세 가지 중심도지표를 사용해 귀농인구 이동 네트워크에서 중요한 역할을 하는 지역을 살펴본 결과 유출입이 모두 많은 지역과 유입과 유출을 동시에 고려한 중요지역을 각각 betweenness점수와 eigenvector점수를 통해 확인했으며 HITS 알고리즘의 hub와 authority점수를 사용하여 지역의 중요도를 반영한 중요도 순위를 알아 보았다.

또한 본격적인 귀농인구 이동 분석을 위해 잠재 변수를 사용한 네트워크 모형을 사용 하였으며 비슷한 유출입 경향을 가진 지역들을 같은 그룹으로 분류하기 위해 stochastic block model을 사용하였다. 그 결과 경상도나 전라도 지방의 군 지역처럼 유입이 활발한 그룹, 광역시처럼 유출이 각각 활발한 그룹, 그리고 각 지방의 시 지역처럼 유출입이 모두 활발한 그룹으로 분류하고 이들 잠재 그룹 간 이동 발생확률을 추정해 귀농인구의 움직임에 지역적 현상이 있음을 확인할 수 있었다. 이 지역적 현상은 특정 지방에서 두드러지는데, 경상도 지방의 군 지역으로 이동한 귀농인구를 살펴보면 타 지방의 대도시에서 유입되는 경우보다 경상도 지방의 대도시에서 유입되는 경우가 훨씬 많았다. 전라도 지방에서도 이와 비슷한 현상이 존재하며 서울과 인천에서 유입되는 귀농인구를 제외하면 두 지방의 지역적 현상은 더욱 두드러져 보였다. 이에 반해 경기도 지방은 전국의 다양한 지역으로 유출되는 경우가 내부에서 이동하는 경우보다 우세하였다.

Stochastic block model은 귀농인구의 이동 경향 확인에는 유용하지만 이동인구의 존재 여부만을 사용하였기 때문에 개별 지역의 특징을 반영할 수 없다. 이를 해결하기 위해 latent distance model을 사용해 추가 분석을 진행한 결과 잠재 공간에서 지역들은 같은 지방에 속한 지역끼리 뭉쳐있는 경향을 확인할 수 있었으며 이는 stochastic block model의 결과와 비슷하였다. 또한 귀농인구 이동의 방향을 설명하기 위해 지역의 귀농인구 유출과 유입 경향을 의미하는 sender effect와 receiver effect 변수를 사용하고 이것을 지역 간 인구수 차이 변수의 효과와 함께 살펴보았다. 그 결과 지역 간 인구수 차이 변수의 포함에 따라 두 random effect의 분산이 감소하는 경향이 있지만 그것만으로는 random effect의 효과를 충분히 설명한다고 할 수 없다. 따라서 귀농인구 이동에는 지역 간 거리와 인구수 차이 외에 다른 요인이 존재한다고 추측할 수 있다. 위의 여러가지 네트워크 모형은 귀농인구 이동 경향을 잘 설명하고 있고 향후 이에 관련된 추가적인 요인의 효과를 분석하는 데 유용한 모형이 될 것이라 예상된다.

References

- Chae, S. H. (2013). Analysis of the current status and success factors for returning farmers with the case

- study of the migration status for returning farmers program graduates, In *Proceedings of the Korean Society for Agricultural Education and Human Resource Development 2013 Meeting*, 43–60.
- Daudin, J. J., Picard, F., and Robin, S. (2008). A mixture model for random graphs, *Statistics and Computing*, **18**, 173–183.
- Davis, K. F., D’Odorico, P., Laio, F., and Ridolfi, L. (2013). Global spatio-temporal patterns in human migration: a complex network perspective, *PLoS One*, **8**.
- Handcock M. S., Raftery, A. E., and Tantrum, J. M. (2007). Model based clustering for social networks, *Journal of the Royal Statistical Society: Series A*, **170**, 301–354.
- Jung, J. H., Roh, J.-S., and Jang, W. (2016). Prediction of rural return migration and analysis of sustainability for returning farmers using big data, *Research Service Report of Rural Development Administration*.
- Jung, J. H., Roh, J.-S., Jang, W., Kim, S. B., Yoon, K. Y., and Kim, J. (2014). Regional analysis of the migration patterns of returning farmers, *Journal of the Korean Society of Rural Planning*, **20**, 221–232.
- Kang, D.-K. (2006). Adoption process with return types of return farmers, *Journal of Korean Agricultural Education*, **38**, 23–53.
- Kim, J. S. (2010). Challenges of sustainable rural development policy with focus on employment and living environment, *Paper presented at the 2014 summer meeting of the Korean Regional Development Association*.
- Kim, J. S. (2014). Social Innovation for rural issues, *Science and Technology Policy*, **24**, 24–36.
- Kleinberg, J. M. (1999). Authoritative sources in a hyperlinked environment, *Journal of the ACM*, **46**, 604–632.
- Krivitsky, P. N. and Handcock, M. S. (2008). Fitting latent cluster models for networks with latentnet, *Journal of Statistical Software*, **24**.
- Krivitsky, P. N., Handcock, M. S., Raftery, A. E., and Hoff, P. D. (2009). Representing degree distributions, clustering, and homophily in social networks with latent cluster random effects models, *Social Networks*, **31**, 204–213.
- Latouche, P., Birmele, E., and Ambroise, C. (2012). Variational Bayesian inference and complexity control for stochastic block models, *Statistical Modelling*, **12**, 93–115.
- Park, M.-H., Lim, J.-E., Seo, H.-S., and Hwang, J.-H. (2014). A study on city people’s returning strategy for economic vitalization of region, *Korea Journal of Agricultural Management and Policy*, **41**, 795–820.
- Roh, J.-S., Jung, J. H., and Jeon, J. Y. (2013). Returning farmers and the aging of farm households: prospects of changes in rural population by their influx, *Journal of the Korean Society of Rural Planning*, **19**, 203–212.
- Shortreed, S. and Handcock, M. S. (2006). Positional estimation within the latent space model for networks, *Methodology*, **2**, 24–33.
- Snijders, T. A. (2011). Statistical models for social networks, *Annual Review of Sociology*, **37**, 131–153.
- Statistics Korea (2014). Year 2013 returning farmers and rural migration, *Press Release*, Retrieved from http://kostat.go.kr/portal/korea/kor_nw/2/7/11/index.board

네트워크 모형을 이용한 귀농인구 이동 분석

이현수^a · 노재선^b · 정진화^b · 장원철^{a,1}

^a서울대학교 통계학과; ^b서울대학교 농경제사회학부, 농업생명과학연구원

(2016년 2월 15일 접수, 2016년 3월 25일 수정, 2016년 3월 30일 채택)

요약

귀농인구의 증가는 빠르게 고령화되고 있는 한국농촌에서 경제 활성화의 단초로 주목받고 있다. 본 논문에서는 2013년 귀농귀촌인 통계를 바탕으로 네트워크 방법을 사용해 귀농인구의 이동에 대한 분석을 진행한다. 구체적으로 다양한 네트워크 중심도지표를 사용해 주요 귀농 거점지역을 파악하고 통계적 네트워크 모형을 사용해 귀농인구의 이동에 존재하는 경향을 살펴본다. 또한 latent distance model을 이용하여 거리가 귀농인구 이동에 미치는 영향에 대해 알아본다.

주요용어: 귀농, 인구이동, 네트워크 중심도지표, HITS algorithm, Stochastic block model, Latent distance model

본 논문은 농촌진흥청 연구사업(주관과제번호: PJ010718)의 지원과 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