Journal of Information Hiding and Multimedia Signal Processing ©2018 ISSN 2073-4212
Ubiquitous International Volume **9**, Number **3**, May **2018**

A Model for Estimating the Correlation of Latent Variable in Social Network and Social Integration

Feng Xiao[1,2], Ling Wu[2], Hao Luo[3,*]
[1]College of management
Heilongjiang University of science and technology, China
[2]College of economics and management
Northeast Agricultural University, China
[3]School of Aeronautics and Astronautics
Zhejiang University, China
*Corresponding author: luohao@zju.edu.cn

Received September, 2017; revised January, 2018

Abstract. *Based on the study of latent variables, this paper uses the estimation model to verify the relationship between the latent variables. Taking the social network and the social integration of rural migrant workers as an example, the quantitative relationship between them is analyzed by mathematical model. We collect relevant data and use exploratory analysis to test the dimensions of social network and social integration of migrant workers. The correlation coefficient analysis shows that the correlation of the social integration dimensions and the social network dimensions is higher.Through regression analysis, we verify the effect of social network on various dimensions of rural migrant workers' social integration. Specially, through logistic regression analysis, we get the data of the influence of migrant workers' social network and demographic variables on social integration intention. The results show that migrant workers' social network and demographic variables have a significant impact on the willingness of social integration. The results of this study can provide a useful reference for expanding the social network of migrant workers and promoting the social integration of migrant workers in China.*
Keywords: Correlation Analysis, Regression Analysis, Factor Analysis, Latent Variable, Social network of Rural Migrant Worker , Social Integration of Rural Migrant Worker

1. **Introduction.** In the correlation analysis of two latent variables, such as the influence social network on social integration of rural migrant workers, we often need to investigate the correlation of social network and social integration. They are all latent variables which are related and decided by one or more observed variables. For example, the social network of rural migrant worker is related to three factors, the social network scale, the social heterogeneity and the social network position. Meanwhile, the social network scale is related to the number of compatriots, the number of friends, the number of persons keeping in touch, and so on. So, the social network is a latent variable. In this paper, we address on estimating the relationships between unobserved constructs (latent variables) from observable variables. We employ Factor analysis model to estimate the dimensions of latent variable. Then the correlation of latent variables is estimated by using Pearson product-moment correlation coefficient and regression analysis.

This paper is organized as followed. In Section 2, we use the regression model to estimate the latent variables. In Section 3, the Pearson product-moment correlation

coefficient is used. In Section 4, we show the model for estimating the correlation of latent variables in social network and social integration.
In Section5, we show the result and give an example analysis by using the dataset of social network and social integration of rural migrant worker. And our method is concluded in Section 6.

2. Using Regression Model to Estimate Latent Variables.

2.1. **Latent variable.** The latent variable reflects two main characteristics. First, it is latent rather than appearing. Secondly, its structure is changeable rather than constant. Finally, latent variables will change in intensity or magnitude. In statistics, latent variables are variables that are not directly observed, and it must be inferred from other variables that are observed. The latent variables usually represent shared variance, or the degree of change between the variables.The latent variables that are not relevant can not produce a potential structural model based on commom factors. The latent variable can be inferred in terms of mathematical models by using observed variables.

Latent variable models are used in a large number of scientific research fields, such as psychology, economics, medicine, physics, artificial intelligence, bioinformatics, agricultural, econometric and management. Probabilistic latent semantic analysis, Factor analysis, Principal component analysis and Analytical Hierarchy Process are always used for inferring latent variables.

2.2. **Logistic regression model.** Logistic regression can be seen as a special case of generalized linear model which measures the relationship between the categorical dependent variable and one or more independent variables by estimating probabilities using a logistic function. The logistic function can be defined as follows [2]:

$$\sigma(\mathrm{t}) = \frac{e^t}{e^t + 1} = \frac{1}{1 + \mathrm{e}^{-t}} \tag{1}$$

We can then express t as follows:

$$\mathrm{t} = \beta_0 + \beta \mathrm{x} \tag{2}$$

And the logistic function can now be written as follows, where F(x) can be interpreted as the probability of the dependent variable [3]:

$$\mathrm{F(x)} = \frac{1}{1 + \mathrm{e}^{-(\beta_0 + \beta \mathrm{x})}} \tag{3}$$

This equation converts a number between -1 and +1 to a probability of [0,1]. In this paper, we define the cost function as a logarithmic loss function:

$$\begin{aligned} &\mathrm{cost}(F(\mathrm{x}),\mathrm{y}) = \text{-ylog}(F(\mathrm{x}))\text{-(1-y)log(1-}F(\mathrm{x})) \\ &= \text{-ylog}[\delta(-(\beta_0 + \beta\mathrm{x}))]\text{-(1-y)log}[1\text{-}\delta(-(\beta_0 + \beta\mathrm{x}))] \end{aligned} \tag{4}$$

Therefore, the loss function is expressed as follows:

$$\mathrm{L}(\beta_0, \beta) = -\frac{1}{m}\sum_{i=1}^{n} \mathrm{cost}(F(\mathrm{x}),\mathrm{y}) \tag{5}$$

Algorithm 1. Estimating the Correlation of the social integration and social network

Input: training data **X**, feature weight β, bias β_0, ranking function **f=logistic(w·x+b)**, iterations **T**, study rate **α**, loss function value of t-th iteration **Lt,** tone **st**

Output: final_β, final_β_0

Initialization: β←0, β_0=0

```
for t = 1 to T
    for x_u in X
        if ((y_u=0 &&f(x_u)>0.5)||(y_u=1&&f(x_u)<0.5))
            β ← β + α*(y^(i) - F(x^(i)))*x^(i)
            β_0 ← β_0 + α*(y^(i) - F(x^(i)))
        endif
        if ((y_u=0 &&f(x_u)<0.5&&f(x_u)>0.5-st)
                ||(y_u=1&&f(x_u)>0.5&&f(x_u)<0.5+st))
            β ← β + α*(y^(i) - F(x^(i)))*x^(i)
            β_0 ← β_0 + α*(y^(i) - F(x^(i)))
        endif
    end for
    calculate Lt
    if isMinimum(Lt)
        final_w←w
        final_b←b
    endif
    if Lt>Lt-1
        w,b←random value
    endif
end for
Output: final_w and final_b
```

2.3. **Parameter Estimation.** The gradient descent method is used to minimize the loss function $L(\beta, \beta_0)$ to evaluate parameter ? and ?0. The gradient of loss function is:

Table 1. Factors of social Integration

Category	Feature	Description
Economic Integration (EI)	x_1	Compared with the local residents, I am satisfied with my income
	x_2	My income can meet me in the local life demand
	x_3	I would like to buy social insurance
	x_4	I would like to buy real estate
Culture Integration (CI)	x_5	I can take the initiative to learn the local language
	x_6	I would like to participate in the local social cultural activity
	x_7	I can accept the local customs
	x_8	I agree with the values of the local
Psychological Integration (PI)	x_9	I think I am a city man
	x_{10}	I feel like I belong to the city
	x_{11}	I wish I could have always been developing in the city
	x_{12}	I would like to associate with urban residents
	x_{13}	The local urban residents are very friendly to me
	x_{14}	The local urban residents are worthy of trust

$$\begin{aligned}
&\frac{\partial}{\partial\beta}(\mathrm{L}(\beta,\beta_0)) = -\frac{1}{m}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\mathrm{cost}(F(\mathrm{x}^{(\mathrm{i})}),\mathrm{y}^{(\mathrm{i})})\\
&= \frac{1}{m}\sum_{i=1}^{n}\frac{\partial}{\partial\beta}\mathrm{y}^{(\mathrm{i})}\log(F(\mathrm{x}^{(\mathrm{i})}))\text{-}(1\text{-}\mathrm{y}^{(\mathrm{i})})\log(1\text{-}F(\mathrm{x}^{(\mathrm{i})}))\\
&= \frac{1}{m}\sum_{i=1}^{n}(\frac{\mathrm{y}^{(\mathrm{i})}}{F(\mathrm{x}^{(\mathrm{i})})})\text{-}\frac{(1\text{-}\mathrm{y}^{(\mathrm{i})})}{1\text{-}F(\mathrm{x}^{(\mathrm{i})})})\frac{\partial}{\partial\beta}F(\mathrm{x}^{(\mathrm{i})})\\
&= \frac{1}{m}\sum_{i=1}^{n}(\frac{\mathrm{y}^{(\mathrm{i})}}{\delta[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]})\text{-}\frac{(1\text{-}\mathrm{y}^{(\mathrm{i})})}{1\text{-}\delta[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]})\frac{\partial}{\partial\beta}\delta[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]\\
&= \frac{1}{m}\sum_{i=1}^{n}(\frac{\mathrm{y}^{(\mathrm{i})}}{\delta[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]})\text{-}\frac{(1\text{-}\mathrm{y}^{(\mathrm{i})})}{1\text{-}[\delta\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]})\delta[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]\!]1\text{-}\delta(\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})}))]\frac{\partial[\text{-}(\beta_0+\beta\mathrm{x}^{(\mathrm{i})})]}{\partial\beta}\\
&= \frac{1}{m}\sum_{i=1}^{n}(\mathrm{y}^{(\mathrm{i})}-F(\mathrm{x}^{(\mathrm{i})}))\mathrm{x}^{(\mathrm{i})}
\end{aligned} \tag{6}$$

$$\frac{\partial}{\partial\beta_0}(\mathrm{L}(\beta,\beta_0)) = \frac{1}{m}\sum_{i=1}^{n}(\mathrm{y}^{(\mathrm{i})}-F(\mathrm{x}^{(\mathrm{i})})) \tag{7}$$

In this paper, we use the random gradient descent method. The update for each sample $(x(i), y(i))$ can be shown as follows:

$$\beta = \beta + \alpha * (y^{(i)}\text{-}F(\mathrm{x}^{(i)})) * \mathrm{x}^{(i)} \tag{8}$$

$$\beta_0 = \beta_0 + \alpha * (y^{(i)}\text{-}F(\mathrm{x}^{(i)})) \tag{9}$$

Table 2. Factors of of social network

Category	Feature	Description
Social Network Scale (SNS)	y_1	I have many fellow-villagers in the city
	y_2	I meet a lot of friends and colleagues
	y_3	I know lots of people
	y_4	I keep in touch with many acquaintances
Social Heterogeneity (SH)	y_5	Jobs I contact with are quite different
	y_6	I can get to know some friends have a certain social status
	y_7	I know some people who are very strong in economy
	y_8	My circle of friends is very wide not just fellow-villagers
	y_9	Someone who is not a relative provides me employment opportunities
	y_{10}	My friends are from different areas
Social Network Position (SNP)	y_{11}	I have a certain appeal in my circle of friends
	y_{12}	Most of my friends are willing to take the initiative to contact me
	y_{13}	I can provide valuable employment information
	y_{14}	People who do not know each other in my circle of friends establish contacts
	y_{15}	I am in the center of my circle of friends

In order to accelerate the speed of learning, we use the TONE technology. The TONE has a parameter a distance interval. The samples in the distance, even if the correct judgment, are needed to learn their features. Algorithm 1 gives the parameter estimation to get the logical regression model accord to Eq. (8), (9).

2.4. **Feature Variables.** For predicting the influence of social network on social integration, we choose fourteen influencing factors which grouped by three categories as the feature variables of social Integration.Table 1 describes the factors of social Integration.

And we design fifteen features to represent the social network shown in Table 2.

2.5. **Estimating the Latent Variables.** By using the influencing factors of social integration as features, we learned three logistic regression models to represent three kinds of social integration latent variables. By using x1 to x_4, we built a regression model to express Economic Integration $EI(x)$ as:

$$x_1 = EI(m) = \frac{1}{1 + \mathrm{e}^{-(\alpha_0+\alpha_1 u_1+\alpha_2 u_2+\alpha_3 u_3+.\alpha_4 u_4)}} \tag{10}$$

For Culture Integration $CI(x)$, we use x_5 to x_8, while for $PsychologicalIntegrationPI(x)$ we employ x_9 to x_{14}.

$$x_2 = CI(m) = \frac{1}{1 + \mathrm{e}^{-(\beta_0+\beta_1 u_5+\beta_2 u_6+\beta_3 u_7+.\beta_4 u_8)}} \tag{11}$$

$$x_3 = PI(m) = \frac{1}{1 + \mathrm{e}^{-(\gamma_0+\gamma_1 u_9+\gamma_2 u_{10}+\gamma_3 u_{11}+\gamma_4 u_{12}+\gamma_5 u_{13}+\gamma_6 u_{14})}} \tag{12}$$

Similarly, we trained three social network models to express the latent variables, $Socialnetworkscale\ SNS(y)$, Social Heterogeneity $SH(y)$, Social Network Position $SNP(y)$ are defined as:

$$y_1 = SNS(y) = \frac{1}{1 + \mathrm{e}^{-(\lambda_0+\lambda_1 v_1+\lambda_2 v_2+\lambda_3 v_3+\lambda_4 v_4)}} \tag{13}$$

$$y_2 = SH(y) = \frac{1}{1 + \mathrm{e}^{-(\varepsilon_0+\varepsilon_1 v_5+\varepsilon_2 v_6+\varepsilon_3 v_7+\varepsilon_4 v_8+\varepsilon_5 v_9+\varepsilon_6 v_{10})}} \tag{14}$$

$$y_3 = SNP(y) = \frac{1}{1 + \mathrm{e}^{-(\omega_0+\omega_1 v_{11}+\omega_2 v_{12}+\omega_3 v_{13}+\omega_4 v_{14}+\omega_5 v_{15})}} \tag{15}$$

3. **Correlation Coefficient.**

3.1. **Definition.**

Correlation analysis can be used to determine whether the two variables are related. At the same time, it provides a tool to measure the linear relationship between variables. In general, correlation coefficient can accurately reflect the correlation degree between variables. The Pearson product-moment correlation coefficient which was developed by Karl Pearson in the 1880s is a measure of the linear correlation between two variables X and Y[4].The value of Pearson product-moment correlation coefficient is between +1 and ?1.Pearson's correlation coefficient is usually represented by the letter r, Supposing that the dataset $x1, ..., xn$ has n value, and another set of data $y1, ..., yn$ with n value, then r can be defined as [5]:

$$r = r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \tag{16}$$

where

$$\begin{aligned} \mathrm{x_i} &= EI(m_i), \\ \mathrm{y_i} &= SNS(z_i)' \\ \overline{x} &= \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{x_i}, \\ \overline{y} &= \frac{1}{\mathrm{n}}\sum_{\mathrm{i=1}}^{\mathrm{n}} \mathrm{y_i} \end{aligned} \tag{17}$$

using the following formula for r:

$$r = r_{xy} = \frac{\sum x_i y_i - n\overline{xy}}{(n-1)s_x s_y} \tag{18}$$

where

$$\mathrm{s_x} = \sqrt{\tfrac{1}{n-1} \sum_{i=1}^{n} (x_i - \overline{x})^2},\ \mathrm{s_y} = \sqrt{\tfrac{1}{n-1} \sum_{i=1}^{n} (y_i - \overline{y})^2} \tag{19}$$

3.2. **Pearson's correlation and least squares regression analysis.** The square of the sample correlation coefficient is typically denoted r^2. If we have the observed dataset $y_1, \ldots, y_n$ and the fitted dataset $f_1, \ldots f_n$, and we denote the fitted dataset $f_1, \ldots f_n$ with $\hat{y}_1, \ldots, \hat{y}_n$, then as a starting point the total variation in the Y_i around their average value can be decomposed as follows:

$$\sum_i (Y_i - \bar{Y})^2 = \sum_i (Y_i - \hat{Y}_i)^2 + \sum_i (\hat{Y}_i - \bar{Y})^2 \tag{20}$$

Next, we apply a property of least square regression models, that the sample covariance between Y_i and $Yi - \hat{Y}_i$ is zero. Thus, the sample correlation coefficient between the observed and fitted response values in the regression can be written[6]:

$$\begin{aligned} r(Y, \hat{Y}) &= \frac{\sum_i (Y_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum_i (Y_i - \bar{Y})^2 \cdot \sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i (Y_i - \hat{Y}_i + \hat{Y}_i - \bar{Y})(\hat{Y}_i - \bar{Y})}{\sqrt{\sum_i (Y_i - \bar{Y})^2 \cdot \sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i [(Y_i - \hat{Y}_i)(\hat{Y}_i - \bar{Y}) + (\hat{Y}_i - \bar{Y})^2]}{\sqrt{\sum_i (Y_i - \bar{Y})^2 \cdot \sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sqrt{\sum_i (Y_i - \bar{Y})^2 \cdot \sum_i (\hat{Y}_i - \bar{Y})^2}} \\ &= \sqrt{\frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2}} \end{aligned} \tag{21}$$

Thus

$$r(Y, \hat{Y})^2 = \frac{\sum_i (\hat{Y}_i - \bar{Y})^2}{\sum_i (Y_i - \bar{Y})^2} \tag{22}$$

4. **Model for Estimating the Correlation.** Figure 1 shows the model for estimating the correlation of latent variables in social network and social integration.

In Figure 1, we divide the social integration features into three categories. The feature u_1 to u_4 are responsible for economic integration, the feature u_5 to u_8 are used to determine the culture integration, and the feature u_9 to u_{14} are exploited to compute the psychological integration. We use the logical model to estimate the three social integration latent variables. Similarly, the feature v_1 to v_4 are responsible for social network scale, the feature v_5 to $v_1 0$ are applied to determine the social heterogeneity and v_{11} to v_{15} are used to estimate the social network position. Then the achieved latent variables are used to relationship computation. Then we can get the co-relationship of each latent variable of social integration and social network.

5. **Example Analysis.** We take social network and social integration of rural migrant worker as the example to verify the proposed method in this Section. In order to testify the proposed method on the research field of the influence of social network on the social integration of rural migrant worker, we design the questionnaires and collect 426 effective

Table 3. Factor load and reliability of migrant workers' social network

	SNS	SH	PI	Communalities	Cronbach's α	
A1	**.852**	.203	.175	.797		
A2	**.834**	.131	.144	.733		
A3	**.833**	.130	.238	.768	0.8378	
A4	**.730**	.203	.182	.695		
A5	.225	**.786**	.124	.684		
A6	.107	**.756**	.147	.605		
A7	.162	**.726**	.151	.576		
A8	.170	**.761**	.148	.574	0.8606	
A9	.148	**.683**	.153	.559		
A10	.153	**.724**	.235	.519		
A11	.235	.180	**.713**	.520		
A12	.239	.105	**.695**	.684		
A13	.146	.229	**.688**	.605	0.7796	0.8915
A14	.133	.159	**.657**	.576		
A15	.441	.140	**.713**	.668		
Feature value	5.116	6.913	4.060			
variance contribution rate（%）	18.036	21.753	16.721	56.51		

Table 4. Factor load and reliability of migrant workers' social integration

	EI	CI	PI	Communalities	Cronbach's α	
B1	**.719**	.126	.108	.713		
B2	**.692**	.097	.157	.698	0.842	
B3	**.659**	.308	.117	.657		
B4	**.652**	.316	.096	.586		
B5	.236	**.852**	.061	.524		
B6	.206	**.819**	.252	.619	0.7556	
B7	.209	**.712**	.212	.829		
B8	.328	**.665**	.061	.740		
B9	.098	.011	**.854**	.699		
B10	.152	.109	**.816**	.635		
B11	.114	.067	**.755**	.809	0.8295	0.8716
B12	.280	.243	**.710**	.734		
B13	.125	.161	**.822**	.710		
B14	.170	.032	**.762**	.613		
Feature value	2.916	2.858	3.160			
variance contribution rate（%）	18.156	18.665	19.759		56.58	

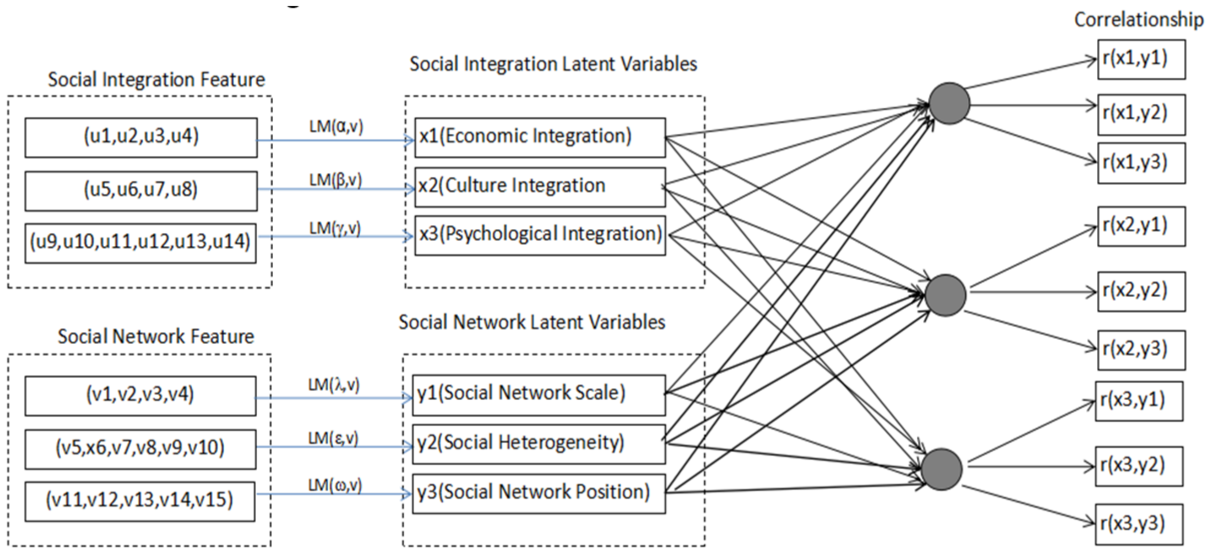

Figure 1. The Model for Estimating the Correlation

Table 5. Correlation analysis of social network and social integration

	Social network scale	Social network heterogeneity	Social Network Position
Economic Integration	0.665**	0.589**	0.519**
Culture Integration	0.604**	0.677**	0.651**
Psychological Integration	0.513**	0.712**	0.704**

results which record the social network and social integration data of rural migrant workers. The questionnaires are all paper questionnaire. Questionnaires were distributed 500, recovery of 489 copies, of which there are 426 effective questionnaires and the effective questionnaire recovery rate of up to 85.2%. The basic characteristics of the respondents shows that the male respondents (61.03%), having not received skills training of migrant workers(66.2%), most of them at the age of 35 years and below (62.2%), the education level of high school is below 89.2%, mainly in the service (40.8%), the manufacturing industry(35.1%), construction (24.1%). On the whole, the sample basically reflects the low level of migrant workers education, most of the new generation of migrant workers, as well as the main characteristics of the service industry and manufacturing enterprises.

5.1. **Factor and Reliability Analysis.** In order to prove that the questionnaire has good reliability, we do the following analysis. Table 4 shows the factor load, reliability of the social network and its each dimension respectively. Table 5 shows the factor load and reliability of migrant workers' social integration. From Table 4 and Table 5, we can see that the total alpha coefficient was 0.89 and 0.87. And Social Network Scale, Social Heterogeneity and Social Network Position coefficient of each dimension is 0.84, 0.86, 0.78. And the Economic Integration, Culture Integration and Physiological coefficient of each dimension is 0.84, 0.76, 0.83. The statistical analysis of the factor analysis and reliability test showed that the social inclusion questionnaire used in this study has good reliability.

5.2. **Correlation analysis of social network and social integration.** Table 6 shows the result of correlation analysis of social network and social integration. Table 6 lists the

Table 6. Correlation analysis of social network and social integration

dependent variable	independent variable	Standard regression coefficient	t	Sig	Adjusted R Square
Economic Integration	Social Network	0.612	21.346	.000	0.377
Economic Integration	Social network scale	0.382	11.046	.000	0.465
	Social heterogeneity	0.418	13.110	.000	
	Social network position	0.367	9.987	.000	
Culture Integration	Social Network	0.587	24.515	0.00	.435
Culture Integration	Social network scale	0.271	9.478	0.00	.512
	Social heterogeneity	0.483	16.916	0.00	
	Social network position	0.374	13.383	0.00	
Psychological Integration	Social Network	0.533	19.913	0.00	0.341
Psychological Integration	Social network scale	0.196	4.478	0.00	0.447
	Social heterogeneity	0.476	11.916	0.00	
	Social network position	0.301	7.334	0.000	

Pearson correlation coefficients among the variables of this research. From the data in the table, we can see that network scale, network heterogeneity and network position are significantly positive correlation with the social integration of rural migrant workers at 0.01 level. The network scale and economic integration are the highest correlation with the coefficient being $0.665 * *$, network heterogeneity and the psychological integration correlation coefficient is $0.712 * *$, and correlation coefficient of social networking position and psychological integration is $0.704 * *$. But the correlation coefficient is lower between network heterogeneity and the psychological integration. Overall, the network scale and economic integration are of the highest correlation.

5.3. **Regression analysis of social network and social integration.** Regression analysis results of social network and each dimension of social integration are showed in Table 6. The experimental results show that when the whole social network is viewed as independent variables, economic integration, cultural integration, psychological integration are viewed as the dependent variable, the standardized regression coefficients are respectively $0.612, 0.587, 0.533$. At the same time, when we take the dimensions of the social network of rural migrant worker as independent variables, and take economic integration, cultural integration, the psychological integration as the dependent variable, the regression equation of F value are are respectively $120.265, 98.365, 89.322$.The coefficient of determination adjusted multiple regression is $0.465, 0.512.0.447$, which show that the social network has a good explanation for social integration of rural migrant workers. From the significance test of regression coefficient, the significant level is all less than 0.05, which pass the test. Therefore, from the perspective of the results, the social network and its dimensions for the positive impact of three dimensions of social integration have been verified Logistic regression analysis results of demographic variables, social network and willingness of social integration are showed in Table7.

6. **Conclusion.** Based on the results of exploratory factor analysis, as the latent variables, social integration has been extracted from three common factors, namely economic integration, cultural integration and psychological integration. Social network has also

Table 7. Logistic regression of social network and willingness of social integration

	B	Wald	Std.Error	Sig	Exp(B)
sex	0.352	3.615	0.209	0.048	1.322
age	0.748	10.037	0.230	0.004	2.233
Education level	0.676	8.124	0.317	0.024	2.089
Skill training	1.102	11.144	0.450	0.011	3.432
Social network scale	1.321	7.987	0.666	0.045	4.67
Social network heterogeneity	2.677	15.341	0.6441	0.009	10.75
Social Network Position	3.156	5.785	0.765	0.016	13.654
Constant	1.597	2.561	0.764	0.092	0.23

been extracted from three common factors, namely network scale, network heterogeneity and network location. Pearson correlation coefficient is used to estimate the correlation degree between variables. The results show that all dimensions of social network have good correlation with all dimensions of social integration. Taking the various dimensions of social networks as independent variables, and taking each dimension of social integration as the dependent variable, we use regression analysis to carry out data processing, the results show that the independent variable has a significant effect on the dependent variable. Using logistic regression analysis to measure the correlation of variables, we come to the conclusion that sex, age, education level, skill training, social network scale, social network heterogeneity and social network position have a positive impact on the desire for social integration of migrant workers.

Acknowledgment. This work is supported by Philosophy and Social Sciences Planning Project of Heilongjiang Province(No.15JYE04), Philosophy and Social Sciences Planning Project of Heilongjiang Province(No.14B042),National Natural Science Foundation of China (No.71341049), Humanities and social sciences research program of the Ministry of Education(No.12YJAZH157).

REFERENCES

[1] D. A. Freedman Statistical models: theory and practice, *Cambridge university press*, 2009.
[2] J. D. W. Hosmer, S. Lemeshow, R. X. Sturdivant, Applied logistic regression, *John Wiley & Sons,* 2013.
[3] J. M. Hilbe, Logistic regression models , *CRC Press*, 2009.
[4] A. J. Onwuegbuzie, L. Daniel, N. L. Leech, Pearson product-moment correlation coefficient, *Encyclopedia of Measurement and Statistics*, vol. 1, pp. 750-755, 2007.
[5] M. T. Puth, M. Neuhuser, G. D.Ruxton, Effective use of Pearson's productmoment correlation coefficient, *Animal Behaviour*, vol. 93, pp. 183-189, 2014.
[6] S. Prion, K. A.Haerling, Making sense of methods and measurement: pearson product-moment correlation coefficient, *Clinical Simulation in Nursing*, vol. 10, pp. 587-588, 2014.
[7] H. C. Huang, The configuration of Social Network Structure and Knowledge Innovation Capability in the Taiwanese Orchid Industry, *Journal of Network Intelligence*, vol. ,3, pp. 250-262, 2017.
[8] M. L. Daniel, G. G. Cristian, G. D. Vicente, B. Cristina Pelayo G-Bustelo, and Nestor Garcia-Fernandez. SenseQ: Replying questions of Social Networks users by using a Wireless Sensor Network based on sensor relationships, . *Data Science and Pattern Recognition*, pp. 1: 1-12, 2017.