Network Centrality and Activities in Small Social Networking Sites (SNS)

Hideki FUJIYAMA (Dokkyo University)
Tatsuhiro SHICHIJO (Osaka Prefecture University)

1. Introduction

Social Networking Sites/services (SNS) have become popular all over the world in the last decade. In Japan, the total number of membership in SNS is seventy one million, as of January 2009. The largest site (Mixi) had a membership of about 15,680,000 as of September 2008. In addition, many other major Social Networking Service Providers exist, such as Facebook, Myspace, Yahoo!Days and so on.

The distinctive function of a Social Networking Sits/services (SNS) allows people to connect with others by mutual consent, as well as acknowledge them as friends. This is important in social relationships in SNS, as such relationships constitute online social networks. In addition, individual activities (such as personal blogs) have externality, as the activity is observed by others (mainly their friends) and constitutes as a main features.

In the last decade, there has been a great interest in the theoretical study of networks. In general, the structure of a social network affects individual behavior (Bramoullé and Kranton (2007)). However, it is very difficult to detect the systematic relationship between social networks and individual behavior because of the multiple equilibria that exist, even in a simple network games (Galeotti et al. (2010)).

Hence, it is important to evaluate empirical studies on network analysis. In Ballester et al. (2006), under the simple but general externality effects that come from the network, they show that the Nash equilibrium action of each player is proportional to the Bonacich centrality. The purpose of this paper is to examine the theoretical relationship between Bonacich centrality and individual behavior using actual data from SNS.

In the following sections, we show that a significant correlation exists between them under a partial adjustment. That is, Bonacich centrality is a significant factor in the explanation of activities in SNS.

2. First Order Condition and Bonacich Centrality

In this section, we review the model and results in Ballester et al. (2006). Player i's payoff function given effort level $x_i \geq 0$ is defined as:

$$u_i(x_1, \ldots, x_n) = \alpha x_i + \frac{1}{2}(-\beta - \gamma)x_i^2 + \sum_{j \neq i}(\lambda g_{ij} - \gamma)x_i x_j \tag{1}$$

where $\alpha > 0,\ \beta > 0,\ \gamma \geq 0,\ \lambda > 0$.

This is known as another expression of bilinear payoffs. In the Social Network Services context, we have the following interpretation:

- α : parameter on effort level, such as writing blog entries, reading other members' blogs, and so on.
- β : some costs for activities in SNS
- γ : global (uniform) substitutability, but it is natural to assume that in SNS this parameter is relatively small. global (uniform) substitutability, but it is natural to assume that in SNS this parameter is relatively small.

- $\lambda \cdot g_{ij}$: network complementarity in SNS, such as writing blogs by one's friends results in more activity for one's own blog.

Given the utility function, Ballester et al. (2006) show that the first-order condition is expressed as:

$$x_i^* = \frac{b_i}{\sum_i b_i} \sum_j x_j^* \tag{2}$$

where b_i is an adjusted Bonacich centrality obtained from original Bonacich centrality by affine transformation. This equation shows that each player's effort depends on her centrality and aggregate effort levels.

3. Regression Model

To obtain a regression model, we introduced additional factors. Because we use monthly data (see section 4), we take into account inertia in the activities in SNS. Hence, under a partial adjustment, lagged dependent variables were included.

This is expressed as:

$$x_{i,t} - (\eta_1 x_{i,t-1} + \eta_2 x_{i,t-2}) = \theta[x_{i,t}^* - (\eta_1 x_{i,t-1} + \eta_2 x_{i,t-2})], \tag{3}$$

where x^* is the theoretical effort level given in section 2, x is the actual effort level, θ is the parameter for speed of adjustment, and η is the parameter for inertia in different periods. By using the first-order condition, we have the following regression model:

$$x_{i,t} = \eta_1(1-\theta)x_{i,t-1} + \eta_2(1-\theta)x_{i,t-2} + \theta BX_{i,t}^* + \mu_i + \varepsilon_{it} \tag{4}$$

where $BX_{i,t}^* \equiv \left(\frac{b_i(\boldsymbol{G},\lambda^*)}{\sum_{j\neq i} b_j(\boldsymbol{G},\lambda^*)}\right) \sum_{j\neq i} x_j^*$, μ_i is individual specific effect, and ε_{it} is the reminder disturbance. It is important to keep in mind that there is a linear restriction on parameters, such that:

$$\eta_1(1-\theta) + \eta_2(1-\theta) + \theta = 1. \tag{5}$$

4. Data

We use two data sets obtained from Social Network Services in the current operation.

The first SNS (SNS1) is mainly for university students and was founded for educational purposes. Students from about seven different universities participate in it, and some universities use it in classes. Members can also use it for ordinary and/or private purposes, such as blogging, email, viewing profiles, and so on.

The second SNS (SNS2) is for ordinary people, and they use it to exchange personal information and information relating to hobbies.

We have data from these two SNS on mutually acknowledged friendships, based on a one-year period from SNS1 and about 2.5 years from SNS2. At the end of the period, there were 290 members in SNS1 and 742 members in SNS2. There were 814 relations with friends in SNS1 and 1013 relations with friends in SNS2.

We use the frequency of writing blog entries per month as a representative effort level (x_i) in SNS because it is the main content for SNS. These data constitute panel data for estimation purposes.

We use the frequency of writing blog per month as the representative effort level (x_i) in SNS because it is the main contents in SNS. The adjusted Bonacich centrality is calculated at the beginning of each month. These data constitutes a panel data for estimation.

5. Results and Conclusion

The estimator from Arellano and Bond (1991) is used because of the need to eliminate the effects of lagged dependent variables and the endogeneity of regressors.

Estimated results in SNS1 are shown in Table 1. With this, we verify empirically the results of Ballester et al. That is, each player' s effort depends on her position in the network structure (centrality measure) and aggregate effort levels (peer effects). The standard test for zero autocorrelation in first-differenced errors was passed and the linear combination constraints on the parameters was not rejected.

As for the test of over-identifying restrictions, we cannot conduct it because, in SNS activity (writing blog entries), the contents and motivations vary greatly. Hence, it is difficult to assume that the disturbance term (ε_{it}) is homoscedastic.

Table 1: Bonacich Centrality and Activity

```
Arellano-Bond dynamic panel-data estimation  Number of obs         =      1245
Group variable: playerid                     Number of groups      =       228
Time variable: t
                                             Obs per group:    min =         1
                                                               avg =  5.460526
                                                               max =         9

Number of instruments =    108               Wald chi2(3)          =     63.30
                                             Prob > chi2           =    0.0000
One-step results
                               (Std. Err. adjusted for clustering on playerid)
------------------------------------------------------------------------------
             |               Robust
      wrifre |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       x(-1) |   .4103557   .1008239     4.07   0.000     .2127444    .6079669
       x(-2) |   .1635145   .0732802     2.23   0.026     .0198879     .307141
        BX^* |   .3658257   .1701496     2.15   0.032     .0323386    .6993129
------------------------------------------------------------------------------
Instruments for differenced equation
        GMM-type: L(2/.).x L(2/.). BX^*
------------------------------------------------------------------------------
------------------------------------------------------------------------------
Arellano-Bond test for              | Test for Linear combination of parameters
zero autocorrelation                |
in first-differenced errors         |
   +-----------------------+        | H0: parameter of x(-1)
   |Order |  z     Prob > z|        |      + parameter of x(-2)
   |------+----------------|        |       + parameter of BX^* =1
   |   1  |-2.7348  0.0062 |        |
   |   2  | 1.4861  0.1373 |        |        chi2(  1) =    0.19
   +-----------------------+        |      Prob > chi2 =    0.6667
    H0: no autocorrelation          |
------------------------------------------------------------------------------
```

As for SNS2, we have similar results at the 10% significance level.

To summarize, under the partial adjustment, Bonacich centrality is a significant factor in the explanation of SNS activities.

References

Arellano, M. and S. Bond (1991) "Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations," *The Review of Economic Studies*, Vol. 58, No. 2, pp. 277–297.

Ballester, C., A. Calvó-Armengol, and Y. Zenou (2006) "Who's who in networks. wanted: the key player," *Econometrica*, pp. 1403–1417.

Bramoullé, Yann and Rachel Kranton (2007) "Public goods in networks," *Journal of Economic Theory*, Vol. 135, No. 1, pp. 478–494, July.

Galeotti, Andrea, Sanjeev Goyal, Matthew O. Jackson, Fernando Vega-Redondo, and Leeat Yariv (2010) "Network Games," *Review of Economic Studies*, Vol. 77, No. 1, pp. 218-244, 01.