Classification of Social Network Sites based on Network Indexes and Communication Patterns

F.Toriumi[1], **I.Okada**[2], **H.Yamamoto**[3], **H.Suwa**[4], **K.Izumi**[5] and **Y.Hashimoto**[5]
Nagoya University, Aichi, Japan[1]
Soka University, Tokyo, Japan[2]
Rissho University, Tokyo, Japan[4]
University of Electro-Communications, Tokyo, Japan[4]
The University of Tokyo, Tokyo, Japan[5]

Abstract

We analyzed data for a large number of small social network services (SNSs) and classified them in terms of their structures and communication patterns. Using this classification, we analyzed their features and found that most of them have small world, scale free, and negative assortativity characteristics. We also classified them on the basis of calculated network indexes and compared the four types. Finally, we classified their communication patterns and identified four types of friend networks: partial, parity, inclusive, and independent.

1 Introduction

As part of the steady growth of new network communication tools, the expansion of social network services (SNSs) such as Facebook and orkut is greatly affecting societies worldwide.

There have been many previous studies of online social networks. Adamic et al. [Adamic *et al.*, 2003], for example, studied a university SNS called Nexus and analyzed its structure and the attributes and personalities of its users. Yuta et al. [Yuta *et al.*, 2007] investigated the network structure of the mixi, and discovered a gap in the community-size distribution that is not observed in real social networks.

Moreover, they developed a simple model to account for this feature. Ahn et al.[Ahn *et al.*, 2007] compared the structures of three online SNSs, (Cyworld, MySpace, and orkut), each with more than 10 million users. They also analyzed the historical evolution of the topological characteristics of Cyworld. These studies mainly focused on large-scale SNSs for general users. In addition to such SNSs, many examples of user-limited SNSs can also be found, such as campus, company, and regional SNSs, that provide specialized services for a limited number of users and thereby effectively stimulate user communication on the Web. These user-limited SNSs are now receiving more attention due to their business potential.

However, SNS studies have been mostly on particular large-scale SNSs, so we cannot say whether their results apply to general features or to special characteristics of SNSs. From the point of view of comparison analysis, a comparison of only a few types of SNS may not produce statistically significant results. We have analyzed a wide variety of SNSs with the aim of classifying them using several approaches.

In this paper, we describe our classification of a large number of small-scale SNSs and our analysis of their features from the viewpoints of network structure and communication pattern.

2 Social Network Data

We analyzed data for 615 SNSs, each with more than 50 users. The data were provided by So-net Entertainment Corporation, which provides SNS support. Using the user relationship data provided, we constructed a friend-network model for each SNS and used it to analyze their network structures.

So-net ’s SNS support has three features in particular.

- Anyone can create a social network service.
- The SNS administrator can choose if the SNS permits “registration” to participate.
- Anyone who registers automatically becomes a friend of the administrator.

The data analyzed included four parameters of particular interest.

1. User(user ID, date on registered)
2. Link(link ID, user id, user ID, date to created)
3. Blog entry(blog ID, user ID, date on entered)
4. Blog comment(comment ID, blog ID, comment user ID, comment date)

3 Network Structure Analysis

3.1 Distribution of network indexes

Previous analysis of the structures of large-scale SNSs, e.g., Cyworld[Ahn *et al.*, 2007] and mixi[Yuta *et al.*, 2007], has shown that SNSs can have both “ small world ” and “ scale free ” characteristics. However, a question remained as to whether these characteristics are commonly found in various sized networks. We thus statistically analyzed data for a large number of SNSs to clarify the characteristics of SNSs.

Average Path Length and Cluster Coefficients

We investigated whether the SNSs we focused on have the small world characteristic by using their average path lengths and cluster coefficients[Watts and Strogatz, 1998].

First, we determined the distribution of average path length L. The average L and standard deviation for the SNSs were respectively 2.13 and 0.339. The mean average path length was approximately 2.1, and about 56% of the average path lengths were between 1.9 and 2.1. That is, SNSs tend to have very short average path lengths.

The average cluster coefficient C and the standard deviation were respectively 0.377 and 0.206. The cluster coefficients had a wide range, $0.1 \leq C \leq 0.8$. However, about 74% of the cluster coefficients were greater than 0.2. That is, SNSs tend to have high cluster coefficients.

These findings indicate that most SNSs have a small world characteristic.

Degree Distribution

We then investigated whether the SNSs have a scale free characteristic by using their degree distributions [Barabási and Albert, 1999].

To determine whether their degree distributions followed a power law, we calculated their determination coefficients, R^2, by using

$$R^2 = 1 - \frac{\sum_i (\log(y_i) - \log(f_i))^2}{\sum_i (\log(y_i) - \overline{\log(y)})^2}, \tag{1}$$

where $\log(y_i)$ is the logarithmically transformed observed values, $\log(f_i)$ is the logarithmically transformed estimated values of the power law obtained by regression analysis of the degree distribution, and $\overline{\log(y)}$ is the logarithmically transformed average of the observed values. This equation shows that the closer the value of R^2 to 1, the closer the distribution follows a power law. The average determination coefficient was 0.631, and the standard deviation was 0.176. This indicates that the degree distributions of SNSs tend to approximately follow a power law.

Next, we calculated the power indexes for the SNSs with a degree distribution that followed the power law, that is, those with R^2 greater than 0.6. There were 409 SNSs that met this condition. The average power index (γ) for these SNSs was -0.908, and the standard deviation was 0.155. The power index for mixi was about -2.4, which is smaller than the SNSs in So-net SNS. Therefore, the ratio of users with many friends was higher than that of mixi.

Assortativity

Finally, we investigated the distribution of the assortativity of the SNSs. Assortativity, r, is defined as an index showing the degree of correlation between connected nodes. Its value range is $-1 \leq r \leq 1$. When the value r is greater, two nodes that both have high degrees tend to be connected. On the other hand, when the value r is smaller a node with high degree and a node with low degree tend to be connected. The calculated r for mixi was 0.121[Yuta *et al.*, 2007], and that for Cyworld was -0.13[Ahn *et al.*, 2007].

The assortativity of our SNSs was -0.471 on average with a standard deviation of 0.207. Interestingly, all of the SNSs except three (0.5% of the number) had a negative assortativity. This indicates that the users with higher degrees tended to connect with users with lower degrees. We partially attribute this to the existence of a core group of members. These core members actively recruit friends to join the network, so they have many links to other members. Moreover, although these friends tend to accept the invitation, only a few become active users. As a result, the active users have higher degrees, and the neighbors have smaller degrees. The assortativities thus become negative.

The results of our network structure analysis are summarized in Table 2.

These network indexes show that the SNSs we investigated have small world, scale free, and negative assortativity characteristics.

Table 1: Network Indexes

	L	C	R^2	γ	r
Average	2.13	0.377	0.631	-0.908	-0.471
Std. Dev.	0.339	0.206	0.176	0.155	0.207

3.2 Comparison with the Other SNSs

We compared the index values we obtained for the So-net SNSs with those for Flickr [Mislove *et al.*, 2007], orkut[Mislove *et al.*, 2007], Cyworld[Ahn *et al.*, 2007], and mixi[Yuta *et al.*, 2007]. As shown in Table 2, these other SNSs respectively had 1,846,198, 3,072,441, 12,048,186, and 363,819 users at the time the data was collected. (The fault data has been omitted.)

As shown in Table 2, average path length, L, power index γ, and assortativity r had various values. The reason for the big difference between the average path length for the So-net SNSs and the other SNSs is attributed to the great difference in the number of nodes. The assortativity is discriminative because the So-net SNSs are leaning a negative direction widely while that of the other SNSs is non-negative. As mentioned, the general network indexes of the So-net SNSs had different values than those of the other SNSs. This suggests that the So-net SNSs have a vastly different friend network structure to other SNSs, while their sites have similar system feature.

Table 2: Comparison of Network Indexes

	No. of Users	L	C	γ	r
Flickr	1846198	5.67	0.313	-1.74	-
orkut	3072441	5.88	0.171	-1.50	-
Cyworld	12048186	-	0.16	-	-0.13
mixi	363819	5.53	0.328	-2.4	0.121
So-net SNSs	257.6	2.13	0.377	-0.908	-0.471

3.3 Classification of SNSs by Clustering approach

We classified the So-net SNSs by using a clustering approach from view point of the network structure, which based on cal-

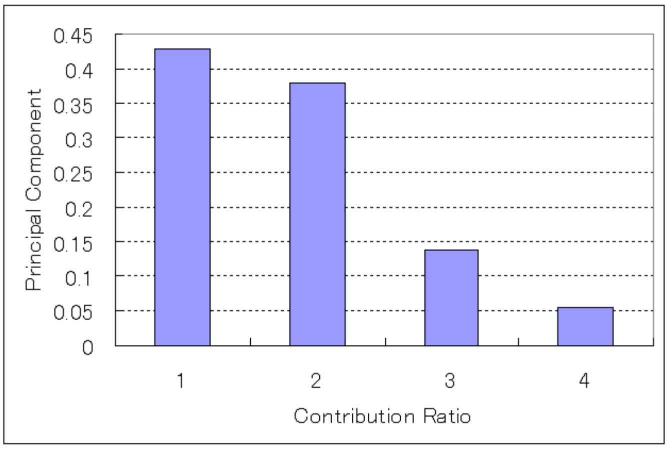

Figure 1: The contribution Ratios of Principal Components

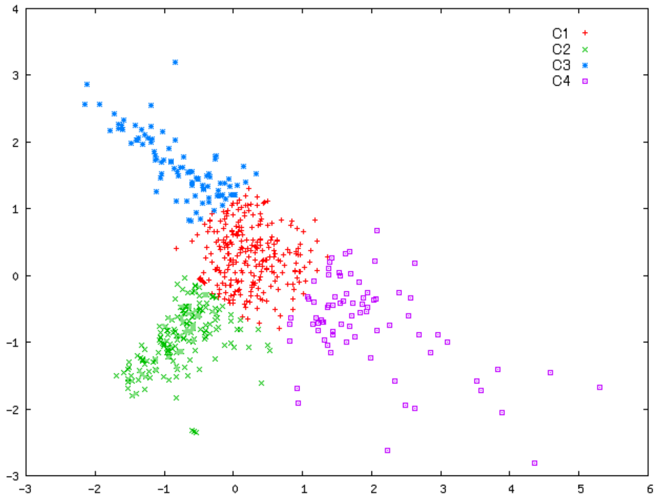

Figure 2: Two-Dimensional Mapping of Network Structures

culated L, C, R^2, and r. Note that γ was not used because some So-net SNSs do not comply with power distribution.

We formulated these four indexes as a character vector v_i,

$$v_i = \left[\frac{L_i}{\sigma_L}, \frac{C_i}{\sigma_C}, \frac{R_i^2}{\sigma_{R^2}}, \frac{r_i}{\sigma_r}\right] \tag{2}$$

where $\sigma_L, \sigma_C, \sigma_{R^2}, \sigma_r$ show standard variations of average path lengths of all SNSs, clustering coefficient, determination coefficients of power low, assortativity, respectively.

In order to observe them easily, we performed principal component analysis, and then clustered the SNSs into four types on the basis of the primary and secondary components because their contribution ratios were notably high. The contribution ratios of principal components are shown in Fig.fig:ContributionRatio A two-dimensional mapping of character vector of each SNSs are shown in Fig.2. We used the k-means clustering method with the number of partitions k equal to 4. To take into account the errors in the initial values produced by this method, we used the case in which the variance ratio among classes was the highest of the multiple cases with different initial values. The calculated average values for the four indexes and the number of SNSs are shown in Table 3 by type. We call each types of SNSs as C1 to C4.

3.4 Characteristics of Clusters

Roughly 40% of the SNSs were classified as C1. These SNSs had both small world and scale free characteristics, so these characteristics should be commonly found in SNSs.

The C2 type SNSs had a substantially smaller average cluster coefficient than the C1 SNSs, meaning that they had a smaller degree of cohesion. Their average assortativity was also substantially smaller, meaning that some of the users in the C2 SNSs exerted traction on the SNS. The average number of users was similar between the two types, but the average number of links in the C2 SNSs was only about 30% that in the C1 SNSs. Moreover, one user in particular had an average of about 78% of the links in the C2 SNSs, meaning that one side of each node pair was almost always the same user. These SNSs thus had an extreme star topology in which one user was connected with all the other users, and all the other users were connected with only that one user.

The C3 SNSs had a higher average cluster coefficient and a shorter average path length L, so they had a higher degree of cohesion. These SNSs had about 88 users on average, which is quite low, but the average degree (number of links) was 16.1, which is extremely high compared to the average for all 615 SNSs (4.94). Therefore, they are close to a complete graph in which members in that SNS are intense relationship.

The C4 SNSs were similar to the C1 SNSs but have longer average path lengths. The average path length for the C4 SNSs was 2.85, significantly higher (0.1% of significant level) compared with that for all the SNSs. Since many of the average path lengths were close to 2.0 because all members tend to connect with the administrator in the So-net SNSs, an average path length greater than 2 means that there are many pairs of users which do not includes administrator. In addition, the C4 SNSs had about 424 friends on average, which is relatively high. This suggests that the C4 SNSs are growing out of the administrator's hands.

The four types of SNSs are diagrammed in Fig.3.

4 Features of SNSs and Analyses of Activation on Users Behaviors

We analyzed the relationship between a user 's friend network and the user 's communication behavior. While a friend network, which consists of links directly, is explicit, communication behaviors is implicit. To determine the correlation between network and behavior, we focused on several features of the communication behavior and identified various patterns of user behavior activation.

We used the 309 Sonet-SNSs sites that satisfied two conditions. 1) The number of users was at least 100 because the analysis would have been meaningless if the number of users actively communicating was small. 2) There were entries of blogs and comments because we needed for our analysis not

Table 3: Average Values of Network Indexes by Four SNS Types

	No. of Users	No. of Links	L	C	R^2	r	No. of SNSs
C1	283.6	1001.1	2.095	0.436	0.713	-0.388	263
C2	236.9	287.5	2.025	0.163	0.573	-0.721	184
C3	87.9	743.0	1.833	0.686	0.380	-0.369	92
C4	423.7	1454.1	2.851	0.313	0.783	-0.280	76

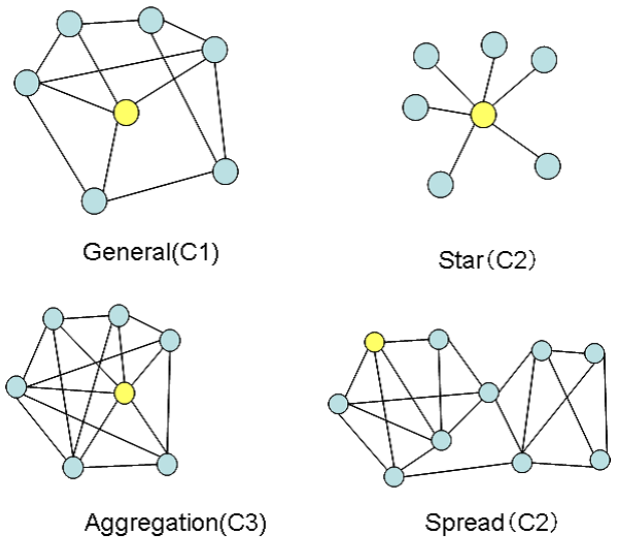

Figure 3: Four Types of SNSs

only the friend network but also blogs and comments to analyze a network of communication behaviors.

We also need appropriate indexes showing how a posted comment on a blog entry is related to the friend network structure, and we need to know what type of behavior patterns the comment has. Can a comment have a “behavior pattern”? To obtain this information, we define an aggregation ratio for friends and a coverage ratio for friends, respectively.

4.1 Index Formulation

The aggregation ratio for friends A is the rato of comments to friends in all the comments. The higher the ratio, the more the comments for blog entries are restricted to friends. The coverage ratio for friends is the ratio of friends who post comments in all the friends who post blog entries. The higher the ratio, the more the actual communications are take place on friend relationships.

$$\text{Aggregation ratio (A)} = \frac{\text{no. of comments for friends}}{\text{no. of all comments}} \tag{3}$$

$$\text{Coverage ratio (C)} = \frac{\text{no. of friends who post comments}}{\text{no. of friends of blog entried user}} \tag{4}$$

4.2 Aggregation and Coverage Ratios

We classified the communication patterns of the SNSs on the basis of the median values of these two indexes (0.737 and 0.610, respectively), as shown in Fig. 4.

- Partial friend network type
  SNSs with a high aggregation ratio and a low coverage ratio. Members communicate with only a limited group of friends.
- Parity friend network type
  SNSs with both high aggregation and coverage ratios. Members communicate within their friend network cyclopaedically, but few communicate with people outside their friend network.
- Inclusive friend network type
  SNSs with high coverage ratio and low aggregation ratio. Members communicate within their friend network cyclopaedically and many communicate with people outside their friend network.
- Independent friend network type
  SNSs with both low aggregation and coverage ratios. Members communications independently of their friend network.

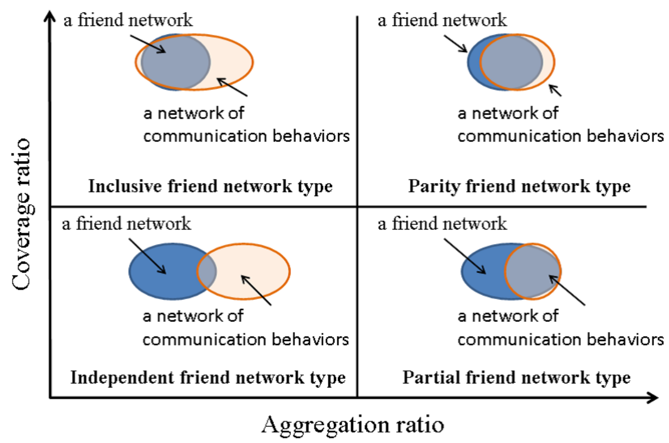

Figure 4: Four Types of Communication Patterns based on Aggregation and Coverage Ratios

4.3 Structural Traits of SNSs based on Communication Pattern

We analyzed several structural traits of SNSs on the basis of their communication patterns: number of users, average degree of cohesion, duration of existence, average path length, cluster coefficient, assortativity, and power index, as shown in

Table 4: SNS Types based on Communication Patterns

communi-cation patterns	N	No. Users	Av. Degree	Duration of Est.	Av. Path Length	Cluster Coef.	Assort-avity	Power Index
Partial	81	131.562	199.599	432.173	2.146	.436	-.360	-.826
Parity	73	137.342	175.342	524.699	2.137	.369	-.403	-.869
Inclusive	81	168.815	128.370	457.728	2.182	.259	-.444	-.940
Independent	74	182.953	115.264	439.014	2.092	.267	-.479	-.935
Chi-Square or F-Value		17.607(x)	45.801(x)	1.795(f)	.946(f)	20.111(f)	4.846(f)	5.525(f)
Significance Prob.		.001***	.000***	.148	.418	.000***	.003**	.001**

$***p < .001$, $**p < .01$

(x) means chi-square value and (f) means F-value.

Table 4. The N in the figure represents the number of SNSs of that type.

The inclusive and independent types have a larger number of users. These types have a low aggregation ratio for friends. This suggests that SNSs with many members who frequently communicate with people outside their friend network tend to be large. The partial and parity types, on the other hand, have high average degrees of cohesion. This suggests that SNSs with many members who limit their communication to within their friend network are thick. The suggestion is supported by the high cluster coefficients of these type SNSs.

4.4 Effect of Communication Pattern on User Behavior Activation

We investigated how the communication pattern affects the activation of user behavior in an SNS. We used the average number of comments posted by a user per day, the number of posting blog entries, the number of user who browsing from PC, and the number of user who browsing from mobile phones as indexes of activation. We tested its differences by using the Kruskal-Wallis test.

As shown in Table. 5, the SNSs with a higher coverage ratio for friends were more active. This makes sense because an SNS is a communication space based on a friend network. A chi-square test showed that the aggregation ratio had no effect on activation. Nevertheless, the communication traits of SNSs do depend on their aggregation ratio. In a parity friend network, communication is only among friends, suggesting that such networks are used to sustain friendships and as a communication tool for everyday matters. In an inclusive friend network, communication is frequently with people outside the friend network, suggesting that they are used mainly to communicate on specific themes or topics.

4.5 Features of Contributive Members and Relationship to Activation

We analyzed the relationship between the activation of user behavior and the contribution of the core users for each type of SNS in order to clarify the role of the core users in the activation. A core user is defined here as a user who plays a central role in network activities such as the administrator. Does the activation pattern when the members on postings are core users different from that when they are edge (not core) users?

To answer this question, we define an index of degree contribution. This index is defined to find whether high-degree user often posts comments or not. The index of degree contribution D_c is calucurated as follows:

$$D_c = \frac{1}{N} \sum_i c_i \cdot d_i \quad (5)$$

where c_i is number of user$_i$'s comments, and d_i is degree of user$_i$. As shown in Table 6, you can observe Kendall's rank correlation coefficients of the index of degree contribution, their significant probabilities, and the indexes of activation for every communicatio pattern.

In the case of the parity friend network, we verify a negative correlation in the contributing degree on links and many indexes on activation and significant tendencies for the number of posting comments and that of posting blog entries. This suggests that there is a negative correlation between the postings of users with a relatively high degree of cohesiveness and activation of behavior in parity type SNSs. The communications in such networks is restricted to within the friend network, and the communication are derived as an extendsion of their daily lives. Therefore, such communication may not need the existence of core members or their involvement.

In the case of the inclusive friend network type, on the other hand, we found a positive correlation relationship between the contributing degree on links and many of the activation indexes. The active involvement of the core members may be needed to activate behavior. Communications tends to go beyond the friend network, so the communications may be for a specific interest or topic rather than for daily matters. These type networks include those for a specific topic such as a disease and those for a specific person, such as a musician. Therefore, the administrator and/or core users play a key role in activating behavior because they work as a traffic controller.

5 Conclusion

We analyzed data for a large number of small SNSs and classified them on the basis of their network structure and their communication pattern. Using the results of this classification, we analyzed several of their features. We found that

Table 5: Effect of Communication Pattern on Activation of User Behavior

Commu-nication pattern	N	Comments No. Postings	Blog Entries No. Postings	From PC No. Browsing	From mobile No. Browsing
Partial Type	81	135.556	145.111	165.654	142.105
Parity Type	73	186.178	170.918	165.849	158.925
Inclusive Type	81	182.296	178.926	160.327	180.432
Independent Type	74	115.649	123.932	126.804	137.405
Chi-square value		34.642	18.066	9.886	11.261
Significant Prob.		.000***	.000***	.020*	.010*

$***p < .001, *p < .05$

Table 6: Relationship between Activation of User Behavior and Contribution of Core Users for Each Type of SNS†

Communi-cation pattern	Comments No. Postings	Blog Entries No Postings	From PC No. Browsing	From mobile No. Browsing
Partial Type	.046	-.057	-.012	-.015
N=81	.541	.448	.870	.845
Parity Type	-.138	-.154	-.129	-.096
N=73	.085+	.054+	.107	.230
Inclusive Type	.180	.137	.174	.052
N=81	.018*	.070+	.021*	.491
Independent Type	.051	.057	.031	-.050
N=74	.517	.469	.692	.526

†Upper values are Kendall's rank correlation coefficients and the lower values are their significant probabilities
$*p < .05$
$+p < .10$

most of them had small world, scale free, and negative assortativity characteristics. We also classified SNSs them on the basis of network indexes and used the results to analyze several other features. A third classification based on communication pattern revealed four types of friend network: partial, parity, inclusive, and independent.

Future work includes clarifying the trajectory of those growing processes by using time analysis. We also plan to analyze the activation and inactivation of behavior by SNS type. The results of the work reported here and of future analysis will enable more effective SNS management.

Acknowledgment

We thank So-net Entertainment Corporation for providing the data of So-net SNS β Version.

References

[Adamic *et al.*, 2003] L.A. Adamic, O. Buyukkokten, and E. Adar. A social network caught in the Web. *First Monday*, 8(6):29, 2003.

[Ahn *et al.*, 2007] Y.Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. *Proceedings of the 16th international conference on World Wide Web*, pages 835–844, 2007.

[Barabási and Albert, 1999] A.L. Barabási and R. Albert. Emergence of Scaling in Random Networks. *Science*, 286(5439):509, 1999.

[Mislove *et al.*, 2007] A. Mislove, M. Marcon, K.P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, pages 29–42. ACM, 2007.

[Watts and Strogatz, 1998] DJ Watts and SH Strogatz. Collective dynamics of'small-world'networks. *Nature*, 393(6684):409–10, 1998.

[Yuta *et al.*, 2007] K. Yuta, N. Ono, and Y. Fujiwara. A Gap in the Community-Size Distribution of a Large-Scale Social Networking Site. *Arxiv preprint physics/0701168*, 2007.