Proceedings of the 2009 IEEE International Conference on Systems, Man, and Cybernetics
San Antonio, TX, USA - October 2009

Human Communication Network Based on the Classification Results of Personal Preferences by Using Self-Organizing Map

Takumi Ichimura*, Azusa Yamasaki†, Akira Hara*, and Tetsuyuki Takahama*
* Graduate School of Information Sciences, Hiroshima City University
3-4-1, Ozuka-higashi, Asaminami-ku, Hiroshima 731-3194 Japan
E-mail : {ichimura, ahara, takahama}@hiroshima-cu.ac.jp
† A graduate of Department of Intelligent Systems, Faculty of Information Sciences, Hiroshima City University

***Abstract*—Social network community may show a kind of properties of complex networks, where key persons called "hub" become bridges to other social network community and then a large network community is formed to combine some small social networks in nature. The relationship between human and human in their network communities was constructed in the virtual space, however, the kind and the strength of their relationships was disregarded. In this paper, based on classified personal preferences by Self-Organizing Map, we propose the adjustment method to decide whether the link between person and person in the network community exists. The initial network formed by BA model and the network changed by our proposed method was analysed by GNS method. We also reported the simulation result for Social Network Services.**

***Index Terms*—Social Network Community, Self-Organizing Maps, Girvan Newman Substructure, Personal Preferences, Social Network Service**

I. Introduction

Social network community may show a kind of properties of complex networks, where some key persons called "hub" become bridges to other networks and then a large community is formed in nature. Recently, Internet community such as Social Network Service (SNS) has been attracted. People participate in some communities and become friends although they has not met mutually, and the size of community has been growing up. A person is regarded as a node in such a community. The nodes and their links form scale-free networks and small-world networks. Yuta tried to extract the Girvan-Newman Substructure (GNS)[2][3] from SNS and investigated the network structure by summarizing the relations among the partial structures[4]. As a result, the extracted GNS was classified roughly into three kinds of sizes; large, medium, and small. They reported that the GNS size skip has been observed. That is, almost medium scale of GNS did not exist in the SNS.

We will make an acquaintance in a human community after we know each other. In such a case, there are some relations among acquaintances who have a characteristic common to hobby and the concern, etc. On the contrary, the network in SNS can make an acquaintance relation easily, even if they don't know each other deeply. We should consider not only the existence of interpersonal relationship but also the strength of the relation in the network community in SNS.

This paper proposes a method of forming social network community based on personal preferences. First, we classify the input-output signals of personal preferences into some categories by Self-Organizing Map (SOM)[1] to investigate how personal preference is similar. Second, we propose some rules of generating/annihilating links among persons in the network. The link is generated when the relation almost equals to the classified category with same personal preferences. If the link among persons having personal preferences classified into the same categories exists and the strength of link is fragile, the link will not keep their relation. In such a case, the link will be annihilated.

In order to verify the effectiveness of our proposed method, we applied it to the network model which is constructed by Barabási and Albert (BA) method[5]. The network structure is extracted by the Girvan-Newman Substructure (GNS) for the BA network model with/without personal preferences. Our proposed method has a parameter of generation/annihilation of link among nodes. Then, we will extract various sizes of GNS according to the parameter. Moreover, we try to apply a Japanese famous SNS and report the results.

II. Self organizing feature map

The basic SOM can be visualized as a sheet-like neural network array as shown in Fig.1. The cells (or nodes) of which become specifically tuned to various input signal patterns or classes of patterns in an orderly fashion. The learning process is competitive and unsupervised, which means that no teacher is required to define the correct output for an input. Only one map node called a winner node at a time is activated corresponding to each input. The map consists of a regular grid of processing units. A model of some multidimensional observations, eventually a vector consisting of features, is associated with each unit. The map attempts to represent all the available observations with optimal accuracy using a restricted set of models. At the same time the models become ordered on the grid so that similar models are close to each other and dissimilar models are far from each other.

A sequential regression process usually carries out fitting to the model vectors. The n is the number of input signals. An

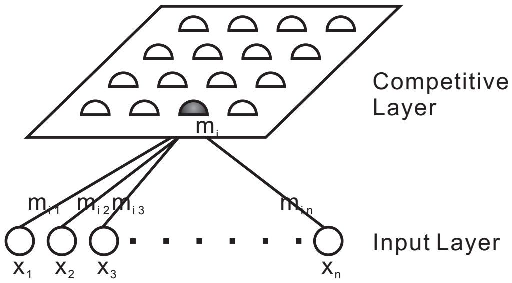

Fig. 1. An overview of SOM

input vector $\boldsymbol{x}$ is compared with all the model vectors $\boldsymbol{m}_i(t)$. The best-match unit on the map is identified. The unit is called the winner. For each sample $\boldsymbol{x} = \{x_1, x_2, \ldots, x_n\}$, first the winner index c (best match) is identified by the condition.

$$\| \boldsymbol{x} - \mathbf{m}_c \| = \min_i \| \boldsymbol{x} - \mathbf{m}_i \| \tag{1}$$

After that, all model vectors or a subset of them that belong to nodes centered around node c are updated at time t as

$$\begin{array}{ll} \mathbf{m}_i(t+1) = \mathbf{m}_i(t) + h_{ci}(\boldsymbol{x}(t) - \mathbf{m}_i(t)) & \textit{for } \forall i \in N_c(t), \\ \mathbf{m}_i(t+1) = \mathbf{m}_i(t) & \textit{otherwise} \end{array} \tag{2}$$

Here h_{ci} is the neighborhood function, a decreasing function of the distance between the ith and cth nodes on map grid. The $N_c(t)$ specifies the neighborhood around the winner in the map array. This regression is usually reiterated over the available samples.

At the beginning of the learning process, the radius of the neighborhood is large and the range of radius becomes small according to the convergence state of learning. That is, as the radius gets smaller, the local correction of the model vectors in the map will be more specific. The h_{ci} also decrease during learning.

III. Analysis of Network Structure

A. Network structure

In network theory, a complex network is a grpah with non-trivial topological features. Two well-known and much studied classes of complex networks are scale-free network and small-world network. Both are characterized by specific structural featurespower-law degree distributions for the former and short path lengths and high clustering for the latter.

A network is named scale-free if its degree distribution, i.e., the probability that a node selected uniformly at random has a certain number of links (degree), follows a particular mathematical function called a power law. The power law implies that the degree distribution of these networks has no characteristic scale.

A network is called a small-world network by analogy with the small-world phenomenon. The small world hypothesis is the idea that two arbitrary people are connected by only six degrees of separation. In 1998, Duncan J. Watts and Steven Strogatz published the first small-world network model, which through a single parameter smoothly interpolates between a random graph to a lattice[6]. Their model demonstrated that with the addition of only a small number of long-range links, a regular graph, in which the diameter is proportional to the size of the network, can be transformed into a "small world" in which the average number of edges between any two vertices is very small while the clustering coefficient stays large.

The BA model is a network model characterized in growth and the preferential attachment for power-law degree distributions. We explain the construction method by BA model in next subsection.

B. BA model

The biggest role of Prof. Barabási has been the introduction of the scale-free network concept in real world network theory. He has studied the growth and preferential attachment, and the mechanism is in part for the structure of the World Wide Web or the cell. The algorithm of BA model is in the following.

The network begins with an initial network of m_0 nodes. It should be noted that $m_0 \geq 2$ and the degree of each node in the initial network should be at least 1, otherwise it will always remain disconnected from the rest of the network.

New nodes are added to the network one at a time. Each new node is connected to m of the existing with a probability that is biased so that it is proportional to the number of links that the existing node already has. Formally, the probability p_i that the new node is connected to node i is given by Eq.(3),

$$p_i = \frac{k_i}{\sum_j k_j}, \tag{3}$$

where k_i is the degree of node i. Heavily linked nodes ("hubs") tend to quickly accumulate even more links, while nodes with only a few links are unlikely to be chosen as the destination for a new link. The new nodes have a "preference" to attach themselves to the already heavily linked nodes. Fig.2 shows the process of growing network by BA model.

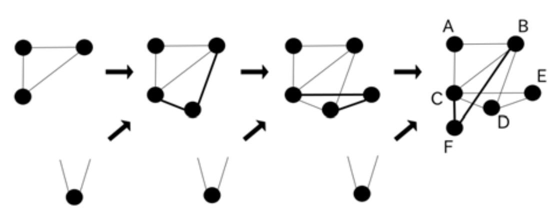

Fig. 2. Growth in BA model

C. Girvan-Newman Substructure

Girvan and Newman [2][3] proposed a computer algorithm based on the iterative removal of edges with high "betweenness" scores that appears to identify such structure with some sensitivity, and this algorithm has been employed by a number of authors in the study of such diverse systems as networks of email messages, social networks, and so on. In this subsection, we explain GNS method briefly.

Let e_{ij} be one-half of the fraction of edges in the network that connect vertices in group i to those in group j, so that the total fraction of such edges is $e_{ij}+e_{ji}$. The only exception will be the diagonal elements e_{ii}, which are equal to the fraction of edges that fall within group i (with no factor of a half). Then $\sum_i e_{ii}$ is the total fraction of edges that fall within groups. All other edges fall between groups. The maximum value of this sum is 1, and a division of the network into communities is good if this quantity is large, meaning it is of order 1. On its own, however, the sum is not a good measure of community structure, since it takes its maximal value of 1 if we put all vertices in a single group together, which is a trivial and not particularly useful form of community structure. A more useful measure of community structure is to calculate the sum $\sum_i e_{ii}$ and then subtract from it the value that it would take if edges were placed at random. Such a measure gives a score of zero to the trivial grouping with only a single community, but nonzero scores to nontrivial groupings. Let a_i be the fraction of all ends of edges that are attached to vertices in group i. a_i is calculated straightforwardly by noting that $a_i = \sum_j e_{ij}$. If the ends of edges are connected together at random, the fraction of the resulting edges that connect vertices within group i is a_i. The modularity is defined as follows.

$$Q = \sum_i (e_{ii} - a_j^i)^2 \tag{4}$$

If a particular division gives no more within community edges than would be expected by random chance, this modularity is $Q = 0$.

Since the joining of a pair of communities between which there are no edges at all can never result in an increase in Q, we need only consider those pairs between which there are edges, of which there will at any time be at most m, where m is again the number of edges in the graph. The change in Q upon joining two communities is given by

$$\Delta Q = e_{ij} + e_{ji} - 2a_i a_j = 2(e_{ii} - a_i a_j) \tag{5}$$

which can clealy be calculated in constant time. The quantities e_{ij} are initially equal to one-half of the corresponding elements of the adjacency matrix of network, i.e., to 1/2 for vertex pairs that are joined by an edge and 0 for those that are not. Following a join, some of the matrix elements e_{ij} must be updated by adding together the rows and columns corresponding to the joined communities.

D. Node links based on personal preferences

This subsection describes the method of investigating whether the node link exists according to personal preferences. First, SOM classifies personal preferences into some categories on the map. The winner nodes on the map are labeled i and the corresponding weight w is $\mathbf{w_i} = (w_{i0}, w_{i1}, \cdots, w_{in})$. D is the norm of $\mathbf{w_i}, \mathbf{w_j} (i \neq j)$, $D = ||\mathbf{w_i} - \mathbf{w_j}||$. Let D_{max} be the maximum of D. For each classified categories, the center of the included weights, G_k, is calculated. According to the difference between G_k and each winner node, the generation and elimination of the node link is determined.

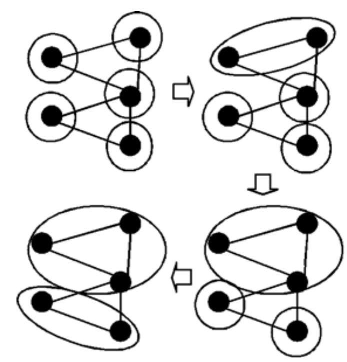

Fig. 3. Identification of Graph Substructure by GNS

The detailed algorihtm is in the following.

- step1: SOM classifies the personal preferences into S categories. Then each node belongs to an arbitrary category in the obtained map.
- step2: For each node $k (k = 0, 1, \cdots, S-1)$, give 2 kinds of threshold values; $Rmin_k$ and $Rmax_k$. $Rmin_k$ is the maximum norm between G_k and each node in the k category. $Rmax_k$ is given by Eq(6). The link between nodes is deleted according to $Rmax_k$.

$$Rmax_k = (1-\gamma) \times (D_{max} - Rmin_k) + Rmin_k \tag{6}$$

  , where γ is a real value in $[0, 1]$ and means the elimination rate of link.
- step3: For all nodes and their links, generate a new link or eliminate the redundant link in the following.
  1) If the norm between G_k and the jth node is less than R_k and if the ith node is not connected to jth node, then generate a new link between their nodes.
  2) If the norm between the ith node in the k category and j node is larger than $Rmax_k$, then annihilate their corresponding link.

The method calculates the degree of similarity for preferences in nodes according to norm of reference vectors in SOM. The longer norm between reference vectors is, the lower degree is. Moreover, if there are links where the personal preference are different from each other, the link will be destroyed. Then, the method deletes some links with low degree of similarity. On the contrary, the group of person who has similar preferences will be constructed, but the personal preference is not unique and consists of some elements. The range of the classified categories is wide and vague. Then, the method calculates the difference between the center of categories is calculated and the node and adds the link according to the difference.

IV. Experimental Results

A. Application to the BA model with 20 nodes

The network is constructed by the BA model, GNS is extracted from the network, and the substructures are investigated. The initial network consists of 4 nodes. An added node to the nework has 3 links to other nodes. BA model stops to add a node till 20 nodes in the network. A circle in Fig.4 shows a node, and the line between nodes shows the link. Fig.4 shows the partial structures by GNS method. Each GNS was denoted by the rectangular in which exist some nodes. Fig.4 in the experimental result shows the 6 kinds of GNS.

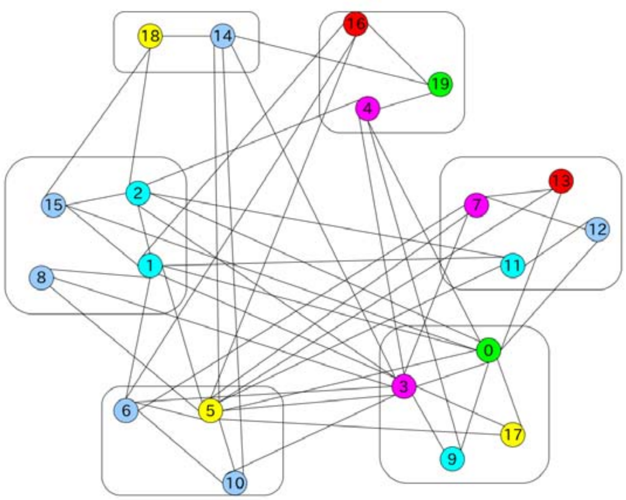

Fig. 4. Network structure with 20 nodes for BA model

Each node in the BA network has attribute values for a personal preference and we classify the values by SOM. In this experimentation, the attribute values 0,1 are given for 10 personal preferences as shown in Table I. The table has no missing data.

Fig. 5 shows the result of classification of attribute values by SOM. The thick line drawn in Fig. 5 means the low degree of similarity of attributes. SOM classified into 6 categories for attributes; ’0’ and ’19’ for category 0, ’1’, ’2’, ’9’, and ’11’ for category 1, ’3’, ’4’, and ’7’ for category 2, ’5’, ’17’, and ’18’ for category 3, ’6’, ’8’, ’10’, ’12’, ’14’, and ’15’ for category 4, and ’13’ and ’16’ for category 5. The attribute values for the personal preferences by using our proposed method described in the subsection III-D was added to the network as shown in Fig. 4. According to the γ in Eq.(6), the obtained substructure by GNS were different. If the γ is small, the number of substructure decreased. If the γ is large, the number of substructure was almost same, but the link between nodes was not easy to be disconnected. The good value of γ is 0.6, when the link was connected according to the attribute values and GNS constructed the substructure to collect nodes in same category. Fig. 6 shows GNS with personal preferences in case of $\gamma = 0.6$.

TABLE I
ATTRIBUTE FOR PREFERENCES

	Attributes									
No	0	1			···					9
0	0	0	0	1	0	0	1	0	0	1
1	0	1	1	1	1	0	1	0	1	1
2	0	1	1	0	0	0	1	0	1	1
3	0	0	0	0	1	0	0	1	0	1
4	0	0	0	0	0	0	0	1	0	1
5	0	1	0	1	1	1	1	0	1	0
6	0	1	1	0	1	1	0	0	0	1
7	0	0	0	1	0	0	0	1	0	0
8	0	0	1	1	0	0	0	1	1	1
9	0	1	1	1	1	0	1	1	1	0
10	0	1	1	0	1	1	1	0	0	0
11	0	1	0	0	0	0	1	1	1	1
12	0	1	1	1	0	1	1	1	0	0
13	0	1	0	0	0	0	0	1	0	0
14	0	0	1	0	0	1	1	0	0	0
15	0	0	1	1	1	1	0	1	1	0
16	0	1	0	0	0	0	0	1	1	0
17	0	0	0	0	1	0	1	0	1	1
18	0	1	0	0	1	0	1	0	1	0
19	0	0	1	0	0	0	0	0	0	1

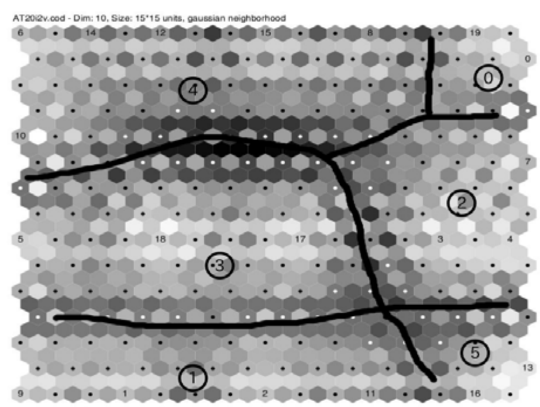

Fig. 5. Classification result by SOM for 20 nodes

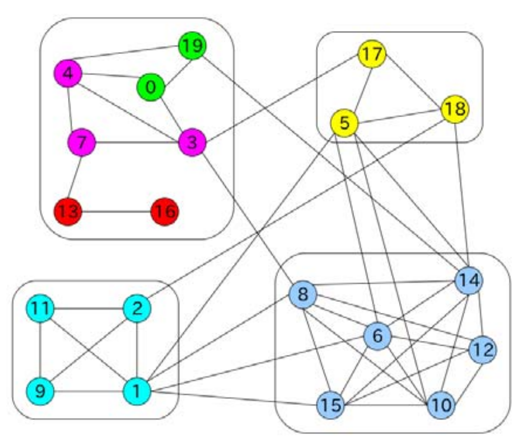

Fig. 6. GNS with personal preferences

B. Application to SNS

This section investigated the relations among 100 users who participated in the community of a certain university in the mixi. 53 people participated to the mixi early and the remaining people participated lately. That is, the relation for not only the current students but also the graduate was constructed. The link of participants in the mixi is shown by the function "MyMixi." The participant selects the personal preferences from the "hobby list." We use the MyMixi and hobby list to investigate the relations among 100 participants. However, MyMixi excluding 100 people was disregarded. Table II shows the 24 kinds of predetermined terms about personal profile in SNS. If the preference term is checked as "yes", the value is 1.

TABLE II
PERSONAL PREFERENCE

Movie watching	Music appreciation	Gourmet	Fashion
Travel	Language Study	TV watching	Gamble
Sports	Karaoke	Drink	Outdoor Sports
Art	Book	Game	Pet
Sport watching	Cooking	Shopping	Drive
Learning	Cartoon	Internet	Beauty and dieting

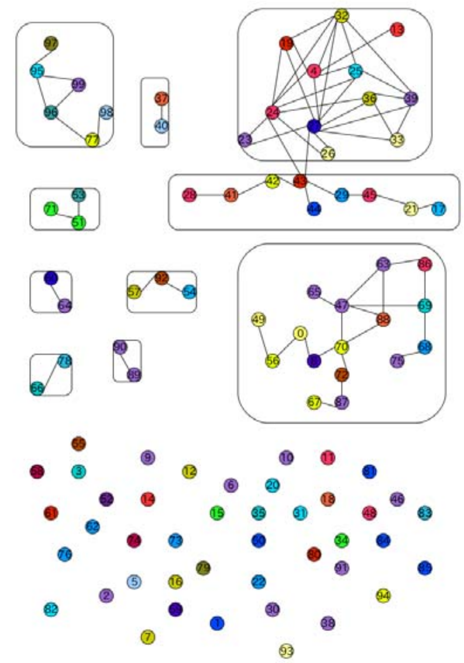

Fig. 7. Identification of Graph Substructure by GNS for SNS

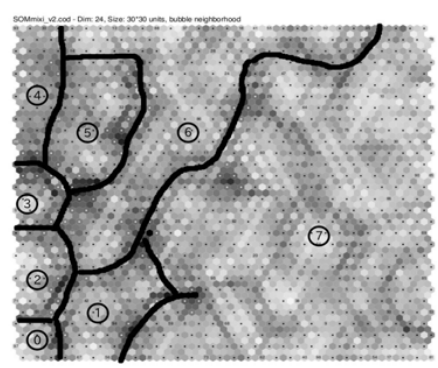

Fig. 8. Classification result for user's preferences

Fig. 7 shows the result of GNS extracting from the mixi network, where a node is the participant and a link is the MyMixi of participants. Many isolated nodes existed at the lower part of figure and the GNS size was also small and the connection between GNSs was hardly seen. This result means that although the connection in a certain grade is seen, the connection between current students and graduate is sparsely. Moreover, the number of cluster is little and there is no small world networks in the GNS. Therefore, the network does not actualize the participant's relation in real world.

Our proposed method can introduce the classification result of personal preferences into the network. First, the attribute values to the data in the "hobby list" was classified by SOM. Fig. 8 shows that SOM classify them into 8 categories on the map. However, category 6 and 7 has many nodes and the accuracy of classification capability is low. Then, the classification by SOM is calculated for category 6 and 7[7]. Fig. 9 and 10 show the classification results for category 6 and 7, respectively.

Fig. 11 shows the result of GNS extracted from the mixi network with personal preferences as shown in Table. II. Although the network community has 43 isolated nodes in Fig. 7, there is only 4 isolated nodes in Fig. 11. The number of isolated nodes has decreased as a result of forming GNS that the exising GNS grows up by connecting the node and the link having similar preferences. Moreover, small-scale GNS has the possibility of growing up to medium-scale GNS by taking the personal preferences in the forming of network community. Therefore, human communication network in SNS requires the connection with personal preferences.

V. CONCLUSION

The method of forming social network community based on personal preference is proposed. SOM classifies the input-output signals of personal preferences into some categories and our method generates and/or annihilates links among persons in the network. The initial network formed by BA model and the network changed by our proposed method was analysed by GNS method. We examined the simulation result for Social Network Services, mixi.

In the GNS extracted from the changed network, the growth of the size was seen compared with former network. We consider that the strength of links in former network structure relates. Although the links are required to keep the partial structure in the network, a weak connection will disappear by taking advantage of the addion of attributes to links. On the contrary, if the attribute values are not high, a strong

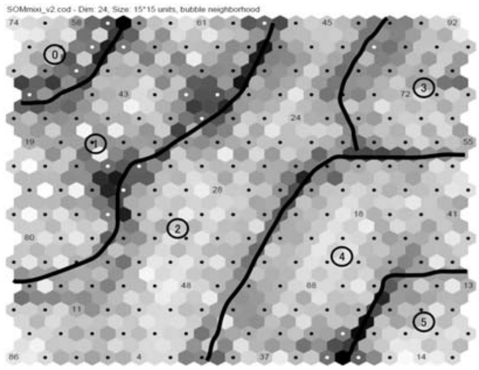

Fig. 9. Reclassification result for the category 6

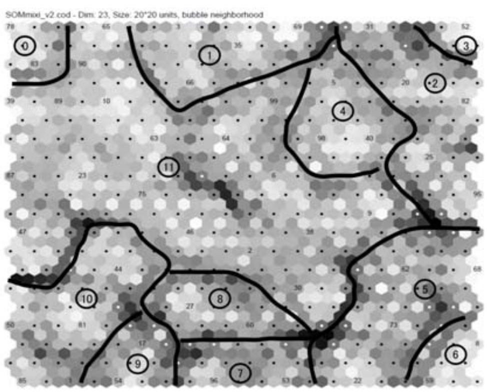

Fig. 10. Reclassification result for the category 7

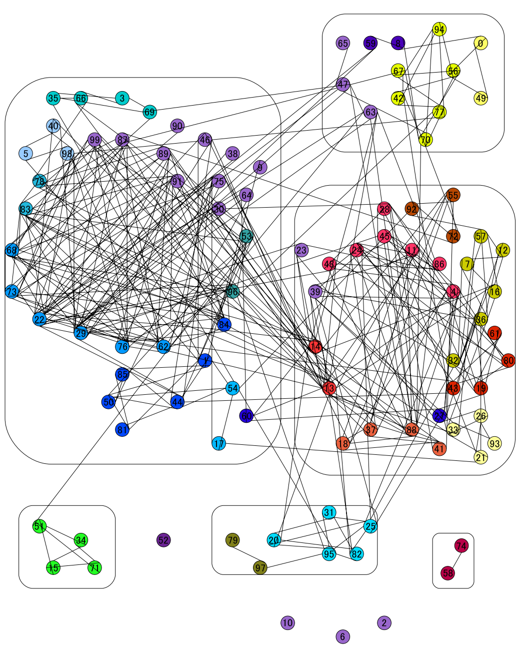

Fig. 11. GNS adopting classification result

connection as shown in "hub" is not demolished and it takes in the other links and grows up to a bigger partial structure.

We will try to give a mathematical proof for an appropriate value of the elimination rate γ. In addition, while SOM classifies the attributes, weight of the attributes and the link for an important item in the hobby list is emphasized. We will propose the method in near future.

REFERENCES

[1] T.Kohonen, *Self-Organizing Maps*, 3rd ed.Springer-Verlag(1997)

[2] M.E.J.Newman, M.Girvan, "Finding and evaluting community structure in networks", PHYSICAL REVIEW E 69, 026113 (2004)

[3] M.E.J.Newman, "Fast algorithm for detecting community structure in networks", PHYSICAL REVIEW E 69, 066133 (2004)

[4] K.Yuta, N.Ono, Y.Fujiwara, "Structural Analysis of Human Network in Social Networking Services(Practical Analysis on Interesting Topics)", Trans. of Information Processing Society of Japan, Vol.47, No.3, pp.865-874(2006)

[5] A.-L.Barabási, R.Albert, "Emergence of scaling in random networks", Science, Vol.286, pp.509-512 (1999)

[6] D.J.Watts, S.H.Strogatz, "Collective dynamics of 'small-world' networks", Nature, Vol.393, pp.440-442 (1998)

[7] T.Yamaguchi, T.Ichimura, K.Mackin, "Adaptive Tree Structured Clustering Method using Self-Organizing Map", Proc. of Joint 4th Intl. Conf. on Soft Computing and Intelligent Systems (SCIS 2008), pp.1412-1417 (2008)