Sensors & Transducers, Vol. 160, Issue 12, December 2013, pp. 423-430

www.sensorsportal.com

Sensors & Transducers

© 2013 by IFSA

http://www.sensorsportal.com

The Acquisition Algorithm for Clustered Subgroup Based on Sensor Balance

[1] Zhang JIANPEI, [1, 2] Li LIHENG, [1] Yang JING, [1] Wang YONG

[1] Computer Science and Technology, Harbin Engineering University, Harbin, 150001, China

[2] Information Center, Heilongjiang University of Chinese Medicine, Harbin, 150040, China

[1] E-mail: zpj323@163.com

Received: 1 November 2013 /Accepted: 22 November 2013 /Published: 30 December 2013

Abstract: One of the main tasks of social aware computing research is crowd behavior and interaction analysis, it is important to understand the formation of community relationship structure. This paper based on perceived balance, proposes the balanced clustered subgroup with its individual coefficients in the environment for group behavior analysis. For transforming iterative optimization in computing process, it also puts forward balance clustering threshold and CIT algorithm for the division of subgroups, and CIOE algorithm for clustering coefficient acquisition. The experimental results show that the algorithm uses the parameters from the integration of time context and social situation on the basis of division in balance clustering subgroup, to acquire individual balance perception values and analyze the aggregation intensity in a group and individual behaviors. *Copyright © 2013 IFSA.*

Keywords: Sensor, Balance clustering subgroup, Balanced relations, Data mining (DM).

1. Introduction

Community is a highly complex social activity involving multiple variables, while group behavior with its interaction analysis is one of the main problems in social perception computing. For instance, through perception computing of social context and time context, it can obtain the balance variation within the community, and interpret the definition of quantized value from a user's perspective, which makes a difference to better understand the evolution of community relations, prejudge the generation of k-faction, k-nuclei and cohesive subgroups, as well as make social situation inference. Hence, the analysis on balance clustering structure of social behavior evolutions in different contexts is one of topic issues in group behavior research.

Previous studies primarily use social network analysis techniques and data mining to analyze the group interactions, including group formation, organizational structure, group activities, communications and its dynamic evolution [2, 3]. The researchers using tripartite groups construct a series of balance theory models, including equilibrium models, clustering models, hierarchical clustering model, and so on. On this basis, using advanced greedy algorithm to generate hierarchical classifications from the entity relationships [7, 8], the basic idea is to determine the hierarchical relationship using pre-defined thresholds based on the entities

similarity. In a further research, it proposes context awareness computing, for instance, using location information, event information and environmental parameters for comprehensive computing to provide users with perceptive information for the analysis of social behavior.

However, most current researches have ignored the perception computing for individual balance clustering based on time and social contexts. For this, this paper presents 2-mode balance clustering subgroups based on temporal extension, and in accordance with subgroup evolution parameters, comprehensive computing and integrating social context and time context, to get the balance clustering coefficient of individuals and subgroups in 2-mode group environment.

2. Research Design

2.1. Theoretical Basis

2.1.1. The Description of Timing-Aware 2-Mode Social Networks

Today, 2-mode network is widespread in our society, such as micro-blog, network evaluation and so on. If uses need for environmental changes of perception group relations, it should analyze the variation of relationship strength. A static community network dataset can be regarded as the temporal snapshot in dynamic evolution for formalized definition of entities and correlation intension based on 2-mode community network.

Definition 1. Timing 2-mode social network G^T: timing community network $G^T = (V^T, E^T, A^T)$ consists of entity set V^T, directed arc set E^T and correlation intensity A^T.

In this case, $T \in (T1 \ldots Tk)$ represents the different time, k as the number of time series; if i and j have correlations at $t1$, then it can be denoted as $e_{i,j}^{t1}$; the initial point of i as $e_{i,j}^{t1}$ is v_i and the terminal point of j as $e_{i,j}^{t1}$ is v_j; the correlation intensity ω from i to j is denoted as ω_e, their vector set as $A^{t1} = [..., \omega_e, ...]^{t1}$; g expresses the connected sub-graph and $g \subseteq G$.

2.1.2. The Balance Clustering of Entity Relationships

Fig. 1 shows the triangular relationships among entities. The ω of entity AB and AC is equal or greater than zero which is expressed by real line, while ω of entity BC is less than zero expressed by dashed line. According to the social network balance theory, supposing Fig.2 shows the entity view of things, where there has the same opinion within AB and AC, and inconsistent in BC. Through ω, it can perceive the imbalance of overall environment among A, B and C.

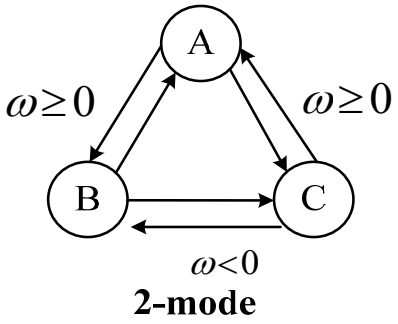

Fig. 1. Triangle relationships.

The balance theory divides the community into two sets, balanced relationship set and unbalanced relationship set; the entity perceives the balanced relationship through the dynamics of ω. However, the real society should classify the entities in community into different forms, which can extends more coherency problems based on the balance theory. The description of clustering subgroup can make the entity perceive the other groups existed in the community and their balanced relationships.

Definition 2. Clustering subgroup g_c: the connected sub-graph g_c constituted by e with $\omega_e \geq 0$ which is within the g_c; e with $\omega_e < 0$ is between g_c, g_c regarded as the clustering subgroup of G.

The division of clustering subgroup belongs to the optimization, which requires repeatedly partition operations, in order to find out the division of minimum arc with $\omega_e \geq 0$ or $\omega_e < 0$ in g_c.

2.1.3. The Balance Clustering in Community Based on Timing Extensions

In order to reduce partition times in the division of community and reflect the balanced clustering changes of entity correlation intensity, it can use setting thresholds. There many measure indexes in social network analysis can be used to set the threshold index, including node degree, betweenness and central potential and so on. However, owing to the different domains, it should choose the appropriate basic indicators in the light of specific problems. This paper mainly focuses on the entity internal changes and correlation intensity variations

between the entities, thus using ω_e as the basic indicator of balance clustering.

Definition 3. Balanced clustering threshold $(P_\omega^{\pm})^T$: taking the average value $P_\omega^{\pm}$ of various timing T correlation intensity with $\omega_e \geq 0$ or $\omega_e < 0$, $P_\omega^{\pm}$ as the balanced clustering threshold of G. The formal description is below:

$$(P_\omega^{+})^T = \{\overline{\sum \omega_e} \mid \omega_e \in A, \omega_e \geq 0\}^T \quad (1)$$

$$(P_\omega^{-})^T = \{\overline{\sum \omega_e} \mid \omega_e \in A, \omega_e < 0\}^T \quad (2)$$

Adopting hierarchical clustering idea and $(P_\omega^{\pm})^T$ to divide the balanced clustering subgroup, presented as below:

Positive-balanced strong clustering subgroup:

$$g_n^{+} = \{(e_{i,j})_n \mid \omega_e \geq P_\omega^{+}, e_{i,j} \in E, \omega_e \in A\} \quad (3)$$

Negative-balanced strong clustering subgroup:

$$g_n^{+} = \{(e_{i,j})_n \mid \omega_e \geq P_\omega^{+}, e_{i,j} \in E, \omega_e \in A\} \quad (4)$$

Positive-balanced weak clustering subgroup:

$$g_n^{-} = \{(e_{i,j})_n \mid \omega_e \leq P_\omega^{-}, e_{i,j} \in E, \omega_e \in A\} \quad (5)$$

Negative-balanced weak clustering subgroup:

$$g_m^{-} = \{(e_{i,j})_m \mid 0 > \omega_e > P_\omega^{-}, e_{i,j} \in E, \omega_e \in A\} \quad (6)$$

From the above divisions, the same entity v at various timing T can be divided simultaneously into the different balanced clustering subgroups. Therefore, it should calculate the balanced clustering coefficient $C(v)$ of entity v to provide users with social relationship intensity variations under the time context. The calculation aims at the proportion of ω_e containing entity v in different $g_{n,m}$; and the superposition of these proportions can obtain the balanced clustering coefficient $C(v)$ of entity v relative to G^T , while $C(v)$ consists of positive-balanced clustering coefficient $C(v^{+})$ and negative-balanced coefficient $C(v^{-})$. Whether $C(v^{+})$ is larger or $C(v^{-})$ is smaller indicates the balanced clustering will be stronger, and vice versa. In a social network based on time series, it mainly through the $C(v)$ variations analyzes the group behavior evolution.

Definition 4. Balanced clustering coefficient with timing extensions $C(v)^T$: supposing the arc set of entity v in timing T respectively as $e(v_n^{+})$, $e(v_n^{-})$, $e(v_m^{+})$ and $e(v_m^{-})$, the related correlation intensity vector set respectively as $\omega_e(e(v_n^{+}))$, $\omega_e(e(v_n^{-}))$, $\omega_e(e(v_m^{+}))$ and $\omega_e(e(v_m^{-}))$, then the $C(v^{+})$ and $C(v^{-})$ as follows:

$$C(v^{+})^T = C(v_n^{+})^T + C(v_m^{+})^T \quad (7)$$

$$C(v_n^{+})^T = \left(\frac{\sum \omega_e(e(v_{n1}^{+}))}{\sum \omega_e(g_{n1}^{+})} + \frac{\sum \omega_e(e(v_{n2}^{+}))}{\sum \omega_e(g_{n2}^{+})} + \ldots + \frac{\sum \omega_e(e(v_n^{+}))}{\sum \omega_e(g_n^{+})}\right)^T \quad (8)$$

$$C(v_m^{+})^T = \left(\frac{\sum \omega_e(e(v_{m1}^{+}))}{\sum \omega_e(g_{m1}^{+})} + \frac{\sum \omega_e(e(v_{m2}^{+}))}{\sum \omega_e(g_{m2}^{+})} + \ldots + \frac{\sum \omega_e(e(v_m^{+}))}{\sum \omega_e(g_m^{+})}\right)^T \quad (9)$$

$$C(v^{-})^T = C(v_n^{-})^T + C(v_m^{-})^T \quad (10)$$

$$C(v_n^{-})^T = \left(\frac{\sum \omega_e(e(v_{n1}^{-}))}{\sum \omega_e(g_{n1}^{-})} + \frac{\sum \omega_e(e(v_{n2}^{-}))}{\sum \omega_e(g_{n2}^{-})} + \ldots + \frac{\sum \omega_e(e(v_n^{-}))}{\sum \omega_e(g_n^{-})}\right)^T \quad (11)$$

$$C(v_m^{-})^T = \left(\frac{\sum \omega_e(e(v_{m1}^{-}))}{\sum \omega_e(g_{m1}^{-})} + \frac{\sum \omega_e(e(v_{m2}^{-}))}{\sum \omega_e(g_{m2}^{-})} + \ldots + \frac{\sum \omega_e(e(v_m^{-}))}{\sum \omega_e(g_m^{-})}\right)^T \quad (12$$

2.2. The Partition Algorithm for Perceived Balance Clustering Subgroup

The algorithm classifies the information in timing 2-mode network context to form high-rise context objects. Fig. 2 shows the calculations of balanced clustering subgroup:

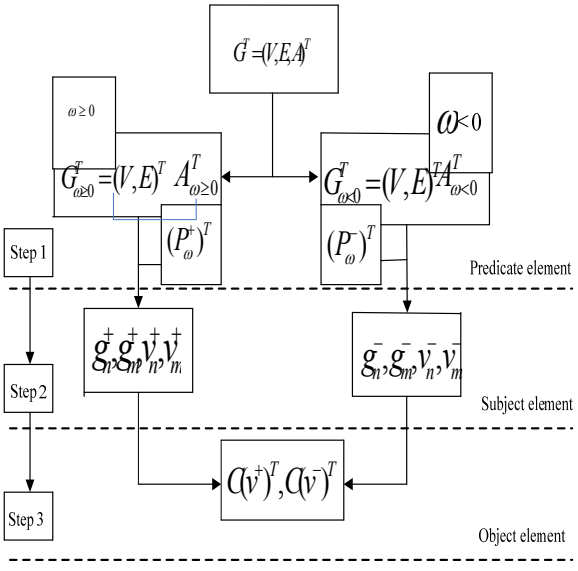

Fig. 2. The acquisition of balanced clustering subgroup with its coefficient.

Step 1: firstly, calculating the context of various timing 2-mode social network, to get $(P_{\omega}^{\pm})^T$;

Step 2: using the obtained $(P_{\omega}^{\pm})^T$ to get balanced clustering subgroup g_n^+, g_n^- and g_m^+ and g_m^-;

Step 3: calculating the timing extension for balanced clustering coefficient $C(v)^T$, to get the balanced clustering subgroup variations based on time context and calculation context.

From the above, step 1 mainly uses formula 1 and 2 to calculate balanced clustering threshold $(P_{\omega}^{\pm})^T$, while step 2 uses $(P_{\omega}^{\pm})^T$ for calculating the balanced clustering subgroup. There is the solution given by CIT algorithm as follows.

Algorithm 1: CIT (Context Interpretation of Threshold value)

Input: G^T

Output: $(P_{\omega}^{\pm})^T$, $(g_n^+, g_n^-, g_m^+, g_m^-)^T$

1 $(P_{\omega}^{+})^T = \{\overline{\sum \omega_e} \mid \omega_e \in A, \omega_e \geq 0\}^T$;

//calculating balanced clustering threshold, formula 1

2 $(P_{\omega}^{-})^T = \{\overline{\sum \omega_e} \mid \omega_e \in A, \omega_e < 0\}^T$;

//calculating balanced clustering threshold, formula2

3 Do Case

4 Case $A(\omega)^T \geq (P_{\omega}^{+})^T$

$(g_n^+)^T = \{(e_{i,j})_n \mid e_{i,j} \in E, \omega_e \in A\}^T$; // getting the arc set of g_n^+

5 Case $0 > A(\omega)^T > (P_{\omega}^{+})^T$

$(g_m^+)^T = \{(e_{i,j})_m \mid e_{i,j} \in E, \omega_e \in A\}^T$; // getting the arc set of g_m^+

6 Case $A(\omega)^T < (P_{\omega}^{-})^T$

$(g_n^-)^T = \{(e_{i,j})_n \mid e_{i,j} \in E, \omega_e \in A\}^T$; // getting the arc set of g_n^-

7 Case $0 > A(\omega)^T > (P_{\omega}^{-})^T$

$(g_m^-)^T = \{(e_{i,j})_m \mid e_{i,j} \in E, \omega_e \in A\}^T$; // getting the arc set of g_m^-

8 End Case;

9 Return $(P_{\omega}^{\pm})^T$, $(g_n^+, g_n^-, g_m^+, g_m^-)^T$;

According to the calculations of balanced clustering subgroup, the CIT algorithm has effectively transformed the iterative optimization of cohesive subgroups algorithm. This is similar to the hierarchical clustering, while the traditional Chameleon algorithm should provide some parameters artificially for selecting K and similarity threshold function, which seems more difficult without priori knowledge. However, CIT algorithm uses the balanced clustering threshold as indicators to get subgroup, effectively transforming the selection of priori parameters and optimization.

Supposing it has M arcs at T, the time complexity of step 1 and step 2 with CIT algorithm is $O(\log M)$, primarily calculating the balanced clustering thresholds. The time complexity from step 3 to step 8 is about $O(2\log M)$, mainly calculating the balanced clustering subgroups. Therefore, the time complexity of CIT algorithm at T is around $O(3\log M)$.

Step 3 mainly uses formula 7 and 10 to calculate balanced clustering coefficients through the balanced clustering subgroups obtained by step 2, and CIOE algorithm presents the solution as below.

CIOE algorithm can figure out the balanced clustering coefficient of every entity in the environment at various moments and observe the coefficient variations, perceive the individual balance evolutions as well as the individual behaviors in the community.

Supposing it has N entities and M arcs at T, the time complexity from step 3 to step 9 is $O(2\log M)$, primarily calculating the sum of relationship strength in every balanced clustering subgroup. The time complexity from step 10 to step 17 is $O(2\log(M+N))$ which is mainly for the balanced clustering coefficient of every entity. So, the maximum time complexity of CIOE algorithm at T is about $O(4\log M \times (M+N))$.

Algorithm 2: CIOE (Context Interpretation of Object Element)

Input: $(g_n^+, g_n^-, g_m^+, g_m^-)^T$

Output: $C(v^+)^T, C(v^-)^T$

1 For $(T_1 \ldots T_k) \in T$

2 For $M_E((g_n^+, g_n^-, g_m^+, g_m^-)^T)$ // M arcs of balanced subgroup

3 Do Case

4 Case $(e \in g_n^+)^T \quad \omega(g_n^+)^T = \omega(g_n^+)^T + \omega_e^T$;

//getting the sum of ω_e with g_n^+

5 Case $(e \in g_m^+)^T \quad \omega(g_m^+)^T = \omega(g_m^+)^T + \omega_e^T$;

// getting the sum of ω_e with g_m^+

6 Case $(e \in g_n^-)^T \quad \omega(g_n^-)^T = \omega(g_n^-)^T + \omega_e^T$;

// getting the sum of ω_e with g_n^-

7 Case $(e \in g_m^-)^T \quad \omega(g_m^-)^T = \omega(g_m^-)^T + \omega_e^T$;

// getting the sum of ω_e with g_m^-

8 End Case;

9 End for;

10 For $N_E(e(v_n^+, v_n^-, v_m^+, v_m^-)^T)$ // N arc sets of entity v

11 Do Case

12 Case $(e(v_n^+) \in g_n^+)^T$

$C(v_n^+)^T = C(v_n^+)^T + (\omega_e(v_n^+) / \omega_e(g_n^+))^T$;

//getting $C(v_n^+)^T$

13 Case $(e(v_m^+) \in g_m^+)^T$

$C(v_m^+)^T = C(v_m^+)^T + (\omega_e(v_m^+) / \omega_e(g_m^+))^T$;

//getting $C(v_m^+)^T$

14 Case $(e(v_n^-) \in g_n^-)^T$

$C(v_n^-)^T = C(v_n^-)^T + (\omega_e(v_n^-) / \omega_e(g_n^-))^T$;

//getting $C(v_n^-)^T$

15 Case $(e(v_m^-) \in g_m^-)^T$

$C(v_m^-)^T = C(v_m^-)^T + (\omega_e(v_m^-) / \omega_e(g_m^-))^T$

//getting $C(v_m^-)^T$

16 End Case;

17 End for;

18 $C(v^+)^T = C(v_n^+)^T + C(v_m^+)^T$;

19 $C(v^-)^T = C(v_n^-)^T + C(v_m^-)^T$;

20 Return $C(v^+)^T, C(v^-)^T$;

21 End for;

3. Empirical Analysis

3.1. Experimental Data and Operating Parameters

Firstly, the experiment uses simulated 2-mode network dataset with 500 nodes, 800 nodes and 1200 nodes for analyzing the execution time of the algorithm. And then, using monk network dataset to analyze balanced clustering subgroup and individual behavior, this dataset includes the emotion data of 25 monks at different timing. The connection between the entities who like each other will be give the values, 3 for the prime choice, then 2 followed by 1. The relations of dislike will be given negative values, -3 preferred and then -2, followed by -1.

The hardware environment in the experiment is Intel Core™ 2 CPU 6320 @ 1.86 GHz 2GB, the operation system Microsoft Windows XP SP3, and to realize under the mixture of Matlab 7.10 and Pajek 2.05. The experimental analysis mainly focuses on operating time and perceived relationships.

3.2. Operating Time Analysis

CIT algorithm is mainly used to calculate balanced clustering subgroups, which is similar to the Chameleon algorithm based on hierarchical clustering. Fig. 3 shows the operating time of these two algorithms in the data environment with 500 nodes, 800 nodes and 1200 nodes. The operating time of both algorithms is added along with the increasing number of nodes, but CIT execution time is lower than Chameleon algorithm. This is because the CIT algorithm uses thresholds for the hierarchy in calculating, decreasing the optimizing frequency.

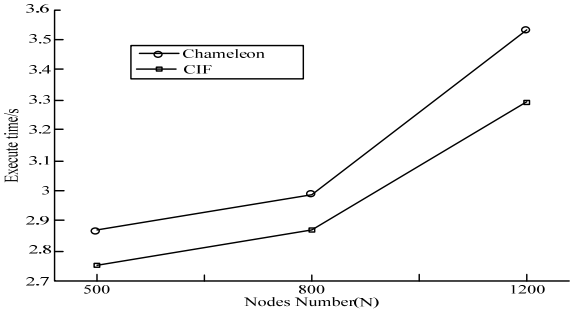

Fig. 3. The execution time of CIT algorithm and Chameleon algorithm.

CIOE algorithm aims at calculating the individual balanced clustering coefficient in the subgroups, so the number of clustering subgroups divided by CIT algorithm using simulation dataset is respectively 53, 69 and 92. Fig. 4 shows the execution time, and the execution time of the algorithm is enlarged along with the increasing number of subgroups, which is related to the number of various subgroups while calculating the time consumption of each entity. As a result, the more the subgroups, the shorter the execution time, and vice versa.

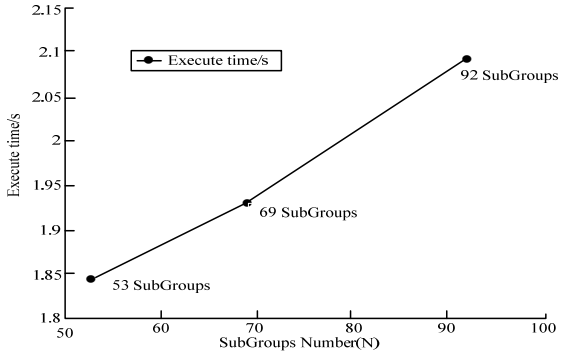

Fig. 4. The execution time of CIOE algorithm.

3.3. Analysis on Balanced Clustering Subgroups with Perceived Relationships

Group clustering research mainly concerns the balance evolution between the entities; when the majority of the relationships have balanced, then resulting in the subgroups. Fig. 5 presents the balanced clustering subgroup topology with five tenses by using CIT algorithm,

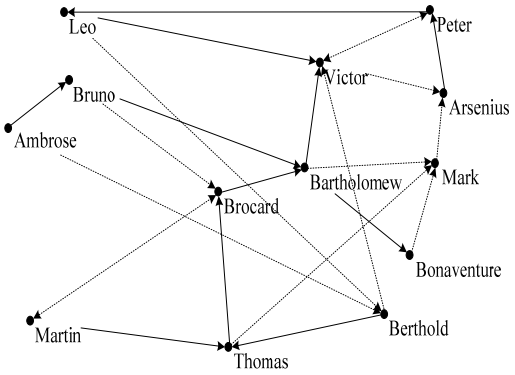

Fig. 5 (a). Balanced clustering subgroups in monk network: g_m^+ and g_m^- at T1.

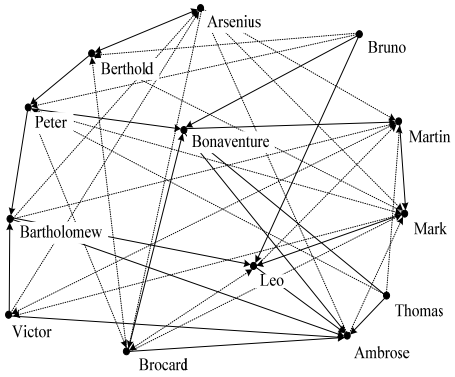

Fig. 5 (b). Balanced clustering subgroups in monk network: g_n^+ and g_n^- at T1.

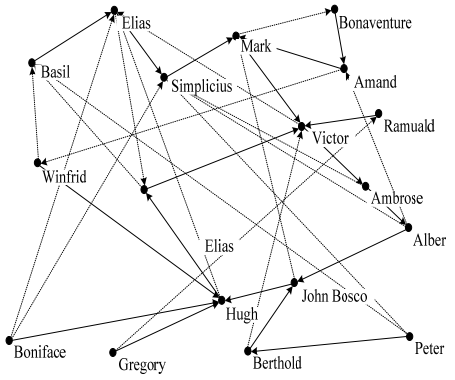

Fig. 5 (c). Balanced clustering subgroups in monk network: c. g_m^+ and g_m^- at T2.

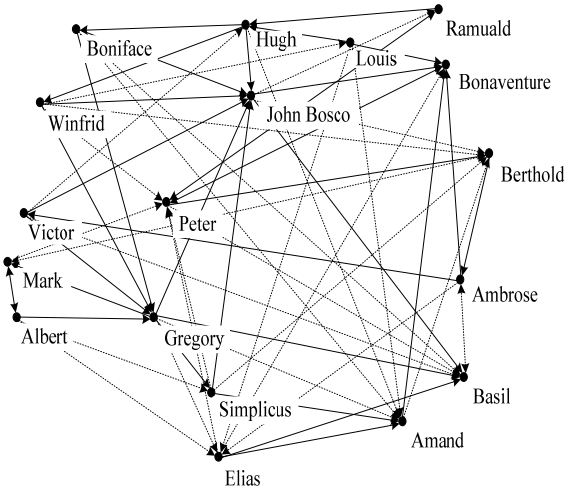

Fig. 5 (d). Balanced clustering subgroups in monk network: g_n^+ and g_n^- at T2.

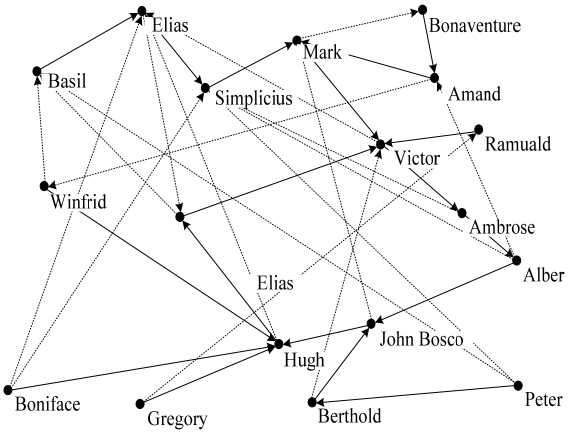

Fig. 5 (e). Balanced clustering subgroups in monk network: g_m^+ and g_m^- at T3.

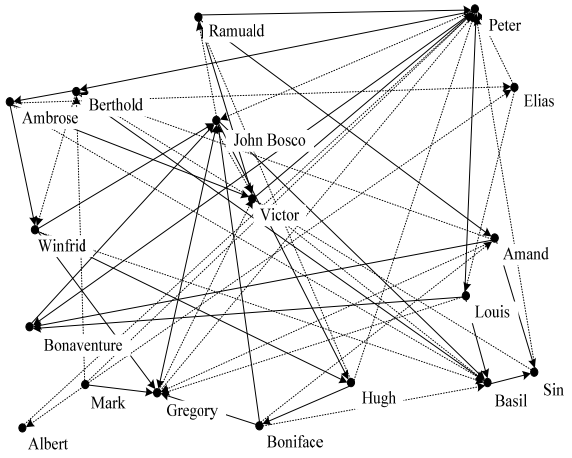

Fig. 5 (f). Balanced clustering subgroups in monk network: g_n^+ and g_n^- at T3.

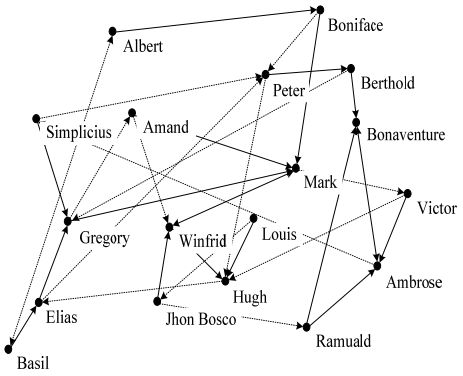

Fig. 5 (g). Balanced clustering subgroups in monk network: g_n^- and g_m^- at T4.

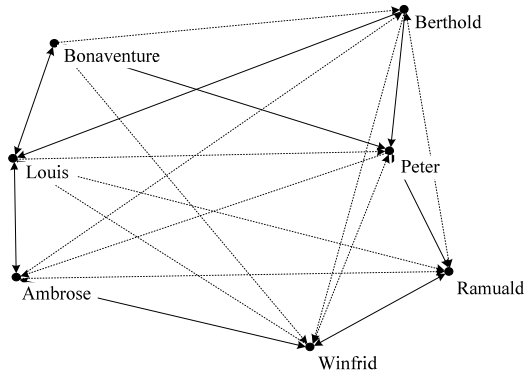

Fig. 5 (j). Balanced clustering subgroups in monk network: g_n^+ and g_n^- at T5.

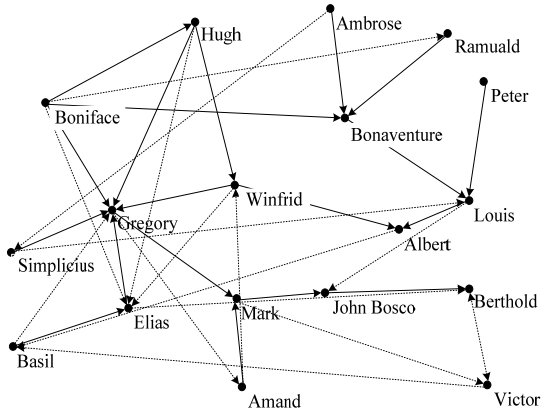

Fig. 5 (h). Balanced clustering subgroups in monk network: g_n^+ and g_n^- at T4.

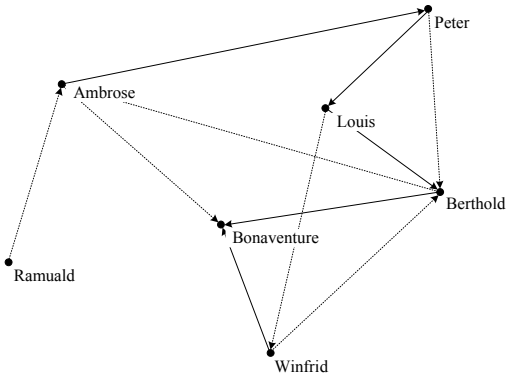

Fig. 5 (i). Balanced clustering subgroups in monk network: g_m^+ and g_m^- at T5.

While the wire of g_n^+ and g_m^+ is the positive value, connected by real line; the wire of g_n^- and g_m^- is the negative value, connected by dashed line. The results show that the same entity has the appearance that doesn't belong to various subgroups at the same time, which indicates the perception for balanced relationships is multiple between the entities, different entity sizes affecting the individual behaviors at the next time interval. For instance, "Martin" exists in different subgroups at T1, while in Fig. 5(a) it selects a relatively small size subgroup only containing three entities, and then "Martin" will eventually leave the monastery with the changing of time; however, Ambrose and Bonaventure practice in the monastery from the beginning to the end.

3.4. Individual Behavior Analysis of Perception Balance

The analysis of perceived balance should have the numeralization for the balance of entities, getting the balanced clustering coefficient by CIT algorithm. Table 1 shows the calculations of partial balanced clustering coefficients in monk network, while the negative-balanced coefficient of Bonaventure is "0" at T3 and T4, illustrating to relationship of dislike has been selected at these two moments. From the timing balanced coefficients of Bonaventure, it can be seen that the positive-balanced coefficients in other moments is greater than negative-balanced coefficients except T2. Thus, it can judge that Bonaventure is harmonious with other monks in the whole community, to be preserved in the monastery.

Table 1. Balanced clustering subgroup coefficients.

Name ID	*T*1		*T*2		*T*3		*T*4		*T*5	
	$C(v^+)$	$C(v^-)$	$C(v^+)$	$C(v^-)$	$C(v^+)$	$C(v^-)$	$C(v^+)$	$C(v^-)$	$C(v^+)$	$C(v^-)$
Martin	0.2462	1.0769	Null	Null	Null	Null	Null	Null	Null	Null
Ambrose	0.2615	0.4615	0.1872	0.1589	0.1254	0.3125	0.7428	0.1402	0.5429	0.5714
Bonaventure	0.3384	0.0769	0.2294	0.5250	0.3603	0.0000	0.7536	0.0000	0.8857	0.4286

Martin left the monastery at T1. It can be known from the Table 1 that Martin has a greater negative-balanced clustering coefficient and distinct, resulting in leaving the monastery.

From the above analysis, it is observed that the balanced clustering subgroups and entity coefficients calculated by CIT and CIOE algorithms can effectively reflect the balance between entities and individual clustering intensity in the community.

4. Conclusions

This paper has discussed the relationship balanced clustering under the time context and social context. Based on the traditional researches, it proposes balanced clustering subgroups and its coefficients based on timing extensions for analyzing group behaviors. It also provides the CIT and CIOE algorithms for calculating clustering subgroups. The algorithms using thresholds effectively decrease the iterations in optimizing, and improve the Chameleon algorithm based on hierarchical clustering through the analysis of execution time. The results show that CIT algorithm can effectively divide the balanced clustering of subgroups, and CIOE algorithm for acquiring the individual balanced clustering coefficient in the community. This not only can provide users with current environmental perception values, but can offer a theoretical basis for the next context inference.

References

[1]. Jin Jie, Analysis on complex network community structure based on relationship strength, *Nanjing University*, 2013.

[2]. Shu-Guang Xiong, Studies on wireless sensor network data collection and query processing, *Harbin Institute of Technology*, 2011.

[3]. Shu-Ling Wang, Studies on resource location in P2P resource sharing system, *China Science & Technology University*, 2012.

[4]. Fang Hong, Hai-Rong Yang, Greedy algorithm and compressed sensing theory, *Acta Automatica Sinica*, Vol. 37, Issue 12, 2011, pp. 1413-1421.

[5]. Si-Hui Tang, Agent-based modeling method for complex social networks and its application, *South China University of Technology*, 2012.

[6]. Hai-Tao Li, Research on position-independent data acquisition in mobile sink sensor networks, *Tianjin University*, 2010.

[7]. Lu Shan, Research on supply chain collaboration based on relational capital and interactive learning, *Central South University*, 2009.

[8]. Zhao Xin, Target depth information acquisition algorithm based on imaging mechanism with its regulating mechanism study, *Yangzhou University*, 2012.

2013 Copyright ©, International Frequency Sensor Association (IFSA). All rights reserved.
(http://www.sensorsportal.com)

Sensors & Transducers Journal (ISSN 1726-5479)

Open access, peer review
international journal devoted to research,
development and applications of sensors,
transducers and sensor systems.
The 2008 e-Impact Factor is 205.767

Published monthly by
International Frequency Sensor Association (IFSA)

Submit your article online:
http://www.sensorsportal.com/HTML/DIGEST/Submition.htm