A New Tripartite Modularity for Detecting Communities

Tsuyoshi Murata

Many users are attracted by online social media such as Delicious and Digg, and they put tags on online resources. Relations among users, tags, and resources are represented as a tripartite network composed of three types of vertices. Detecting communities (densely connected subnetworks) from such tripartite networks is important for finding similar users, tags, and resources. For unipartite networks, several attempts have been made for detecting communities, and one of the popular approaches is to optimize modularity, a measurement for evaluating the goodness of network divisions. Modularity for bipartite networks is proposed by Barber, Guimera, Murata and Suzuki. However, as far as the author knows, there is few attempt for defining modularity for tripartite networks. This paper defines a new tripartite modularity which indicates the correspondence between communities of three vertex types. By optimizing the value of our tripartite modularity, better community structures can be detected from synthetic tripartite networks.

1 Introduction

Relations among real-world entities are often represented as n-partite networks that are composed of n types of vertices. Paper-author networks and event-attendee networks are the examples of bipartite networks, and user-resource-tag networks of social tagging systems are the examples of tripartite networks. Detecting communities (subnetworks that are densely connected inside and sparsely connected outside) from such n-partite networks is practically important for finding similar entities and understanding the structure of social media (Fig. 1).

As a naive approach for transforming n-partite networks into unipartite networks, projection is often employed for the sake of convenience[18]. Suppose $x \in X$ and $y \in Y$ are two types of vertices in n-partite networks. Projection is a transformation from $x_j - y_i - x_k$ connection into $x_j - x_k$ connection so that a network composed of only X-vertices is obtained. However, such projection often loses information that are contained in original n-partite networks. It is pointed out that qualities of the communities obtained from projected networks are worse than those from original non-unipartite networks [9]. This is because projection will generate too dense unipartite networks if the original n-partite networks contain a few high-degree vertices. Another weakness of projection is that there are several ways to project a n-partite network to unipartite networks.

As a metric for evaluating the goodness of detected communities, Newman-Girvan modularity [15] is often employed. Optimizing the modular-
ity is one of the popular strategies for detecting communities from networks. There are some attempts for extending the definition of Newman-Girvan modularity. A modularity for directed networks is proposed by Leicht and Newman [10]. Arenas et al. propose a modularity for weighted directed networks [1].

Since it is not suitable for n-partite networks, some researchers extend its definition for bipartite networks, such as the definitions given by Barber [2], Guimera [9], Murata [12] and Suzuki [16]. As an attempt to extend modularity for tripartite networks, Neubauer recently proposes a tripartite modularity [13] based on Murata’s bipartite modularity. His approach is to project a tripartite network into three bipartite networks and then apply Murata’s bipartite modularity. As far as the author knows, this is the only attempt for defining tripartite modularity. However, Neubauer’s tripartite modularity still needs projection, and projection will lose some of the information that original tripartite network has.

Defining suitable tripartite modularity and optimizing it for detecting communities are practically important for the networks of social tagging systems, which are composed of users, resources and tags. This paper proposes another tripartite modularity for tripartite networks. Experimental results show that our tripartite modularity has abilities of detecting better communities in sparse tripartite networks, and it clarifies the correspondence between communities of different vertex types.

2 Related Works

2.1 Research on Heterogeneous Networks

Many of the previous research on heterogeneous networks employ projection for the sake of convenience. For example, Cattuto et al. [4] project tripartite networks into unipartite networks and analyze their properties.

Ghosh [8] handles heterogeneous networks which are more general than n-partite networks in a sense that any connection between any types of vertices are allowed. Her paper, however, focuses on centralities of such heterogeneous networks.

2.2 Modularity

There is a large number of possible divisions for a network. It is therefore necessary to establish which partitions exhibit a real community structure. Newman-Girvan modularity [15] is a quantitative measurement for the quality of a particular division of unipartite network. Let us consider a particular division of a network into k communities. Let us suppose M is the number of edges in a network; that V is a set of all vertices in the network; and that \( V_1 \) and \( V_m \) are the communities. \( A(i, j) \) is an adjacency matrix of the network whose \( (i, j) \) element is equal to 1 if there is an edge between vertices \( i \) and \( j \), and is equal to 0 otherwise.

Then we can define \( e_{lm} \), the fraction of all edges in the network that connect vertices in community \( l \) to vertices in community \( m \):

\[
e_{lm} = \frac{1}{2M} \sum_{i \in V_l} \sum_{j \in V_m} A(i, j)
\]

We further define a \( k \times k \) symmetric matrix \( E \) composed of \( e_{lm} \) as its \((l, m)\) element, and its row sums \( a_l \):

\[
a_l = \sum_m e_{lm} = \frac{1}{2M} \sum_{i \in V_l} \sum_{j \in V} A(i, j)
\]

If a network had edges between vertices regardless of the communities they belong to, we would have \( e_{lm} = a_la_m \) for this network. Newman-Girvan modularity is thus defined as follows:

\[
Q = \sum_l (e_{ll} - a_l^2)
\]

Newman-Girvan modularity measures the fraction of the edges in the network that connect vertices of the same community minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices. If the number of edges inside communities is no better than random, or if the whole network is taken as a single community, we will get \( Q = 0 \). Values approaching the maximum \( (Q = 1) \) indicate strong community structures.

Newman-Girvan modularity has been employed as the quality function in many community detection algorithms. In addition, modularity optimization is itself a popular method for community detection. There are many related works regarding modularity. Clauset et al. [5] propose a fast mod-

2.3 Bipartite Modularity

2.3.1 Guimera’s Bipartite Modularity

A community is characterized by larger density of intracommunity edges than that of intercommunity edges. However, bipartite networks are different from unipartite networks in that vertices of the same type are not directly connected. For this reason, density of intracommunity edges has to be redefined for bipartite networks. Guimera’s bipartite modularity [9] is defined as the cumulative deviation from the random expectation of the number of Y-vertices in which two vertices of type X are expected to be together:

$$M_B = \sum_{a=1}^{N_M} \left\{ \frac{\sum_{i \neq j} c_{ij}}{\sum_{a} m_a (m_a - 1)} - \frac{\sum_{i \neq j} t_i t_j}{(\sum_{a} m_a)^2} \right\}$$

(4)

where $s$ is a X-vertex community; $N_M$ is the number of X-vertex communities; $a$ is a Y-vertex; $m_a$ is the number of edges that are connected to $a$; $c_{ij}$ is the number of the Y-vertex communities in which vertices $i$ and $j$ are connected; and $t_i$ and $t_j$ are the total numbers of Y-vertex communities to which vertices $i$ and $j$ are connected, respectively.

As you can see, two vertex types are not treated symmetrically in the definition above. Guimera’s bipartite modularity focuses on the connectivities of only one vertex type (via the vertices of the other type). It is therefore not sufficient for representing the connectivities of the other vertex type, which can be defined as follows:

$$M_B' = \sum_{a=1}^{N_M} \left\{ \frac{\sum_{i \neq j} c_{ij}}{\sum_{s} m_s (m_s - 1)} - \frac{\sum_{i \neq j} t_i t_j}{(\sum_{s} m_s)^2} \right\}$$

(5)

In order to measure the connectivities of both vertex types, both $M_B$ and $M_B'$ have to be used.

2.3.2 Barber’s Bipartite Modularity

Modularity is a deviation from a null model, and bipartite networks have specific constraints that should be reflected in the null model. Barber [2] takes the constraints into consideration and formalizes the following bipartite modularity. Since there is no edge between the vertices of the same type, the adjacency matrix of a bipartite network is as follows:

$$A = \begin{bmatrix} 0_{p \times q} & \tilde{A}_{p \times q} \\ (\tilde{A}^T)_{q \times p} & 0_{q \times q} \end{bmatrix}$$

(6)

where $0_{i \times j}$ is the all-zero matrix with $i$ rows and $j$ columns. The probabilities in the null model of an edge existing between vertices $i$ and $j$ are represented as follows:

$$P = \begin{bmatrix} 0_{p \times q} & \tilde{P}_{p \times q} \\ (\tilde{P}^T)_{q \times p} & 0_{q \times q} \end{bmatrix}.$$  

(7)

Barber’s bipartite modularity is defined as follows:

$$Q = \frac{1}{m} \sum_{i=1}^{p} \sum_{j=1}^{q} (\tilde{A}_{ij} - \tilde{P}_{ij}) \delta(g_i, g_{j+p}),$$

(8)

where $g_i$ is the community that vertex $i$ is assigned to, and $\delta(i, j)$ is the Kronecker’s delta. Barber’s bipartite modularity is based on the idea that there is a one-to-one correspondence between the X-vertex communities and the Y-vertex communities. $\delta(g_i, g_{j+p})$ is equal to 1 when the X-vertex community $g_i$ corresponds to the Y-vertex community $g_{j+p}$. This definition implicitly indicates that the numbers of communities of both types are equal. In order to optimize the bipartite modularity, repetitive bipartitioning is employed. Since the number of communities have to be specified in advance, search for an appropriate number of communities is required.

The weaknesses of Barber’s bipartite modularity are: 1) the number of communities have to be searched in advance; and 2) the numbers of communities of both vertex types have to be equal. Both weaknesses come from the bipartitioning method he employs.

2.3.3 Murata’s Bipartite Modularity

In order to overcome the weaknesses of previous bipartite modularities, the constraint of one-to-one correspondence between communities of both types is removed in Murata’s bipartite modularity [12]. One X-vertex community may corresponds to more than one Y-vertex communities and vice versa.

Let us suppose that $M$ is the number of edges in a bipartite network, and that $V$ is a set of all vertices in the bipartite network. Consider a particu-
lar division of the bipartite network into X-vertex communities and Y-vertex communities, and the numbers of the communities are \( L^X \) and \( L^Y \), respectively. \( V^X \) and \( V^Y \) are the sets of the communities of X-vertices and Y-vertices, and \( V^X_i \) and \( V^Y_m \) are the individual communities that belong to the sets \( V^X = \{ V^X_1, ..., V^X_{L^X} \}, V^Y = \{ V^Y_1, ..., V^Y_{L^Y} \} \). \( A(i, j) \) is an adjacency matrix of the network whose \((i, j)\) element is equal to 1 if vertices \( i \) and \( j \) are connected, and is equal to 0 otherwise.

Under the condition that the vertices of \( V_1 \) and \( V_m \) are of different types (which means \( (V_1 \subset V^X \land V_m \subset V^Y) \lor (V_1 \subset V^Y \land V_m \subset V^X) \)), we can define \( e_{lm} \) (the fraction of all edges that connect vertices in \( V_1 \) to vertices in \( V_m \)) and \( a_l \) (its row sums) just the same as those in section 2.2.

\[
e_{lm} = \frac{1}{2M} \sum_{i \in V_1} \sum_{j \in V_m} A(i, j) \tag{9}
\]

\[
a_l = \sum_m e_{lm} = \frac{1}{2M} \sum_{i \in V_1} \sum_{j \in V} A(i, j) \tag{10}
\]

As in the case of unipartite networks, if edge connections are made at random, we would have \( e_{lm} = a_l a_m \). Murata’s bipartite modularity \( Q_M \) is defined as follows:

\[
Q_M = \sum_l Q_{M_l} = \sum_l (e_{lm} - a_l a_m),
\]

\[
m = \text{argmax}(e_{lk}) \tag{11}
\]

\( Q_{M_l} \) means the deviation of the number of edges that connect \( l \)-th X-vertex community and its corresponding \((m\)-th\) Y-vertex community, from the expected number of randomly-connected edges. A larger \( Q_{M_l} \) value means stronger correspondence between \( l \)-th community and its corresponding \((m\)-th\) community.

2.3.4 Suzuki’s Bipartite Modularity

As another attempt for defining modularity for bipartite networks, Suzuki et al. [16] modify Murata’s bipartite modularity. The coefficients of \( e_{lm} \) are changed from \( 2M \) to \( M \), and an average of the connection to all other communities is employed, while Murata’s bipartite modularity chooses corresponding community as the one with the largest number of connection by argmax.

3 Tripartite Modularity

3.1 Neubauer’s Tripartite Modularity

Neubauer [13] proposes a tripartite modularity based on Murata’s bipartite modularity. His approach is to project a tripartite network to three bipartite networks and apply Murata’s bipartite modularity to these bipartite networks.

Let \( g_D, g_U \) and \( g_T \) be the individual communities for each vertex type. Let \( g_X \) and \( g_Y \) be the combination of two communities, assigning elements from domain X or Y to the community given by \( g_X \) or \( g_Y \), respectively. Then Neubauer’s tripartite modularity \( Q_{3B} \) of \((g_D, g_U, g_T)\) with regard to a hypergraph \( H \) is defined as follows, where \( Q_M \) is Murata’s bipartite modularity. \( DU(H) \) is a bipartite network generated by projecting \( H \) to the domains \( D \) and \( U \). \( DT(H) \) and \( UT(H) \) are defined in the same manner.

\[
Q_{3B} = \frac{1}{3}(Q_M(DU(H), g_D \cup g_U) + Q_M(DT(H), g_D \cup g_T) + Q_M(UT(H), g_U \cup g_T)) \tag{12}
\]

The advantage of Neubauer’s tripartite modularity is that it utilizes Murata’s bipartite modularity, while its disadvantage is that projection to bipartite networks causes loss of information that an original tripartite network has.

3.2 Our New Tripartite Modularity

Neubauer points out three properties that are required for \( n \)-partite modularity [13].

**Connectivity** Modularity rewards adjacent vertices being in the same community. In \( n \)-partite graphs, vertices from the same domain by definition are never adjacent. A generalized modularity measure must be able to account for this situation as to avoid punishing the grouping of related, but necessarily non-adjacent vertices.

**Community structure** In \( n \)-partite graphs, different domains may have different community structures. Several communities in one domain may be distinct yet all be connected to one larger community in another domain. A community model of \( k \)-hypergraphs needs to support \( k \) different set of communities, and “being in the same community” needs to be generalized to a more abstract concept of “be-
ing in corresponding communities”.

**Hyper-Incidence** In (non-hyper-)graphs, we only need to consider two vertices per edges. These two vertices are either in corresponding communities or not. With more than two vertices, the binary concept of an edge “connecting two vertices from corresponding communities” or not needs to be generalized to a function describing the degree of correspondence between the vertices’ communities.

According to Neubauer’s paper, his tripartite modularity satisfies the first and the second properties, but the third property is not satisfied since projection is still employed for converting a tripartite network to three bipartite networks. In order to satisfy the third property, a new tripartite modularity should be defined in a way that degree of the correspondence of the communities of three vertex types are clearly indicated.

In addition to that, the following requirements for usual modularity should also be satisfied:

- If all vertices of each vertex type are taken as a single community, its modularity is zero.
- If the number of within-community edges is no greater than the expected value of the same quantity in a network with the same community divisions but random connections between the vertices. In the case of unipartite network, modularity measures the fraction of the edges in the network that connect vertices in the same community minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices.
- Higher modularity value indicates strong community structure.

Let us suppose that a tripartite network \(G\) is described as \((V, E)\), where \(V\) is a set of vertices, and \(E\) is a set of hyperedges. \(V\) is composed of three types of vertices: \(V^X, V^Y, \) and \(V^Z\). A hyperedge connects triples of the vertices \((i, j, k)\), where \(i \in V^X, j \in V^Y, \) and \(k \in V^Z\), respectively. Suppose that \(deg(i)\) is the number of hyperedges that connect to vertex \(i\). The followings are the constraints of a tripartite network:

- The number of edges from \(V^X, V^Y, \) and \(V^Z\) are equal.
  \[
  \sum_{i \in V^X} deg(i) = \sum_{j \in V^Y} deg(j) = \sum_{k \in V^Z} deg(k)
  \]
- \(V\) is composed of disjoint sets of vertices. \(V = \{V^X, V^Y, V^Z\}\), where \(V^X \cap V^Y = V^Y \cap V^Z = V^Z \cap V^X = \emptyset\)

\(A(i, j, k)\) is an adjacency matrix for a tripartite network. The element \(A(i, j, k)\) of the adjacency matrix is 1 if vertices \(i, j, \) and \(k\) are connected with a hyperedge, otherwise it is 0. A community in a tripartite network is defined as a subset of vertices of a single type in this paper, although Barber defines it as a subset of all types of vertices. We employ the above definition since there can be one-to-many correspondence among the communities of different vertex types.

In the case of unipartite network, modularity measures the fraction of the edges in the network that connect vertices in the same community minus the expected value of the same quantity in a network with the same community divisions but random connections between the vertices. In the case of bipartite and tripartite networks, there is no connection between the vertices of the same type. Therefore, our tripartite modularity measures the fraction of the edges in the tripartite network that connect vertices of the corresponding X-vertex communities, Y-vertex communities and Z-vertex communities minus the expected value of the same quantity in a tripartite network with the same community divisions but random connections among X-vertices, Y-vertices and Z-vertices.

Our definition of tripartite modularity is as follows. Let us suppose that \(M\) is the number of hyperedges in a tripartite network, and that \(V\) is a set of all vertices in the tripartite network. Consider a particular division of the tripartite network into X-vertex communities, Y-vertex communities, and Z-vertex communities, and the numbers of the communities are \(L^X, L^Y, \) and \(L^Z\), respectively. \(V^X, V^Y, \) and \(V^Z\) are the sets of the communities of X-vertices, Y-vertices, and Z-vertices, and \(V^X, V^Y, \) and \(V^Z\) are the individual communities that belong to the sets \(\{V^X = \{V^X_1, ..., V^X_L\}, V^Y = \{V^Y_1, ..., V^Y_L\}, V^Z = \{V^Z_1, ..., V^Z_L\}\}\)

\(E^{XY}, E^{YZ}, \) and \(E^{ZX}\) are the sets of the edges that connect vertex pairs \(X\) and \(Y\), \((Y \) and \(Z), \) and \((Z \) and \(X)\), respectively. The number of edges in these sets are equal. \(|E^{XY}| = |E^{YZ}| = |E^{ZX}|\)

Under the condition that the vertices of \(V^X, V^Y, \) and \(V^Z\) are of different types, we can define \(e_{lmn}\) (the fraction of all edges that connect vertices in \(V^X, V^Y, \) and \(V^Z\)) and its sums over three dimensions, such as \(a_l, a_m, \) and \(a_n\).

\[
e_{lmn} = \frac{1}{M} \sum_{i \in V^X} \sum_{j \in V^Y} \sum_{k \in V^Z} A(i, j, k) \quad (13)
\]

\[
a_l^X = \sum_m \sum_n e_{lmn}
\]
argmax if hyperedge connections are made at random,

1. As in the case of unipartite networks, 

\[ Q > 0 \] 

The sum over all communities

\[ n \]

Suppose \( s_X = \sum_l a_X^l, s_Y = \sum_m a_Y^m, \) and \( s_Z = \sum_n a_Z^n. \) From the above definitions, it is obvious that \( s_X = s_Y = s_Z = \sum_m \sum_n e_{lmn} = 1. \) As in the case of unipartite networks, if hyperedge connections are made at random, we would have \( e_{lmn} = a_X^l a_Y^m a_Z^n. \) Therefore, \( Q_X = \sum_m \sum_n (e_{lmn} - a_X^l a_Y^m a_Z^n), \) where \( m, n = \arg\max(e_{ijk}), \) will be zero. On the other hand, if hyperedges from X-vertices are mainly from the vertices in community \( V_X^l, \) the value of \( Q_X \) will be greater than zero. The sum over all communities of \( V^X \) is as follows.

\[
Q_X = \sum_l Q_X^l
= \sum_l \sum_m \sum_n (e_{lmn} - a_X^l a_m a_n) \tag{17}
\]

\( Q_X \) means the deviation of the number of hyperedges that connect \( l \)-th X-vertex community and the corresponding (\( m \)-th) Y-vertex community and (\( n \)-th) Z-vertex community, from the expected number of randomly-connected hyperedges. A larger \( Q_X \) value means stronger correspondence from the \( l \)-th community to the \( m \)-th Y-vertex community and the \( n \)-th Z-vertex community. \( Q_Y \) and \( Q_Z \) are defined in the same manner.

\[
Q_Y = \sum_m Q_Y^m
= \sum_m \sum_l \sum_n (e_{lmn} - a_l a_m a_n) \tag{18}
\]

\[
Q_Z = \sum_n Q_Z^n
= \sum_n \sum_l \sum_m (e_{lmn} - a_l a_m a_n) \tag{19}
\]

Our new tripartite modularity \( Q \) is defined as the average of \( Q^X, Q^Y \) and \( Q^Z. \)

\[
Q = \frac{1}{3}(Q^X + Q^Y + Q^Z) \tag{20}
\]

The main advantages of our new tripartite modularity over Neubauer’s modularity are as follows:

- Our tripartite modularity does not employ projection. As mentioned previously, projection will lose information that original tripartite networks have.

- Our tripartite modularity can be extended to n-partite modularity. Neubauer’s approach of projecting tripartite networks to bipartite ones is not appropriate for general n-partite networks.

4 Experiments

In order to compare our tripartite modularity and Neubauer’s tripartite modularity, experiments are performed using synthetic tripartite networks. Each of the tripartite network for the experiments is composed of 36 vertices (12 vertices each for vertex type X, Y, and Z). The numbers of hyperedges are 10, 20, 30, 40 and 50, and the numbers of communities in each vertex type are 2, 3 and 4. Each X-vertex, Y-vertex, and Z-vertex is assigned to one of these ‘correct’ communities. Vertices of corresponding X-vertex, Y-vertex, and Z-vertex communities are basically interconnected with random hyperedges. The goal of the experiments is to detect ‘correct’ communities from the synthetic tripartite networks by optimizing tripartite modularities.

In order to evaluate the goodness of detected communities, the following normalized mutual information (NMI, \( I(A, B) \)) is employed. This criterion is employed in the experiments for comparing modularities performed by Danon et al. [6],

\[
I(A, B) = -\frac{2 \sum_{i=1}^{c_A} \sum_{j=1}^{c_B} N_{ij} \log \left( \frac{N_{ij} N}{N_A N_B} \right)}{\sum_{i=1}^{c_A} N_i \log \left( \frac{N_i}{N_A} \right) + \sum_{j=1}^{c_B} N_j \log \left( \frac{N_j}{N_B} \right)}
\]

where the number of correct communities is denoted \( c_A \) and the number of detected communities is denoted \( c_B. \) \( N \) is a confusion matrix whose rows correspond to the correct communities and columns correspond to the detected communities. The element of \( N \ (N_{ij}) \) is the number of vertices in the correct community \( i \) that appear in the detected
community $j$. The sum over row $i$ of matrix $N_{ij}$ is denoted $N_i$, and the sum over column $j$ is denoted $N_j$. If the detected communities are identical to the correct communities, then $I(A, B)$ takes its maximum value of 1. If the detected communities are totally independent of the correct communities, $I(A, B) = 0$. Detailed descriptions of the procedures are as follows.

1. Set the number of hyperedges as $M(=10, 20, 30, 40$ or $50)$, the number of vertices as $|V^X| = |V^Y| = |V^Z| = 12$, and the number of communities for each vertex type as $L$ ($= L^X = L^Y = L^Z = 2, 3$ or $4$).

2. Assign $|V^X|/L$ vertices to each of the following 'correct' communities: $V^X_1, \ldots, V^X_L, V^Y_1, \ldots, V^Y_L, V^Z_1, \ldots, V^Z_L$.

3. For each hyperedge $p (1 \leq p \leq M)$, vertices $i(\in V^X_{(p/M)L})$, $j(\in V^Y_{(p/M)L})$ and $k(\in V^Z_{(p/M)L})$ are selected randomly from each community and connected with the hyperedge.

4. After a tripartite network is generated, information about the above communities is hidden. Then communities are detected from the network by optimizing tripartite modularities.

5. Normalized mutual information (NMI) is calculated from detected communities and 'correct' communities.

6. The above procedure is repeated ten times for each parameter and the average of NMI values is computed.

Table 1 shows the results of the experiments. Rows mean the results of optimizing Murata's tripartite modularity and Neubauer’s tripartite modularity, respectively. Columns mean the numbers of 'correct communities'.

Table 1 NMI of Detected Communities

<table>
<thead>
<tr>
<th># of edges = 10</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata</td>
<td>0.442</td>
<td>0.611</td>
<td>0.846</td>
</tr>
<tr>
<td>Neubauer</td>
<td>0.386</td>
<td>0.611</td>
<td>0.846</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of edges = 20</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata</td>
<td>0.428</td>
<td>0.803</td>
<td>0.901</td>
</tr>
<tr>
<td>Neubauer</td>
<td>0.397</td>
<td>0.822</td>
<td>0.940</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of edges = 30</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata</td>
<td>0.525</td>
<td>0.749</td>
<td>0.921</td>
</tr>
<tr>
<td>Neubauer</td>
<td>0.513</td>
<td>0.948</td>
<td>0.946</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of edges = 40</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata</td>
<td>0.604</td>
<td>0.824</td>
<td>1.000</td>
</tr>
<tr>
<td>Neubauer</td>
<td>0.531</td>
<td>0.910</td>
<td>1.000</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th># of edges = 50</th>
<th>2</th>
<th>3</th>
<th>4</th>
</tr>
</thead>
<tbody>
<tr>
<td>Murata</td>
<td>0.815</td>
<td>0.971</td>
<td>1.000</td>
</tr>
<tr>
<td>Neubauer</td>
<td>0.680</td>
<td>0.984</td>
<td>1.000</td>
</tr>
</tbody>
</table>

The optimization of Neubauer’s tripartite modularity is comparable to the optimization of ours for such dense networks, many real heterogeneous networks are sparse. For example, Leskovec et al. analyze about 100 real networks [11]. Densities of these networks vary from 201.78 to 1.25, but their average is 6.69 because most of the networks are sparse. As an example of tripartite networks, Wetzker et al. crawl and analyze the network of Delicious, one of the famous social bookmarking systems [17]. The density of the bookmarks per URLs is 2.62. Based on these facts, we can claim that many real world networks are sparse. The advantage for sparse networks is important for practical applications of real world networks.

5 Conclusion

This paper proposes a new modularity for tripartite networks. Our new tripartite modularity has the following advantages over Neubauer’s tripartite modularity: 1) qualities of detected communities based on optimization of our tripartite modularity are better since projection of tripartite networks are not used for our tripartite modularity, and 2)
our tripartite modularity is extendable for n-partite modularity in general. Our tripartite modularity proposed in this paper is the first step for processing real heterogeneous networks that are available in the Web, such as social tagging systems. Experiments using real tripartite networks and speedup for the calculation of our tripartite modularity are left for our future work.

References