Modularity for Heterogeneous Networks

Tsuyoshi Murata
Tokyo Institute of Technology
W8-59 2-12-1 Ookayama, Meguro
Tokyo, 152-8552 Japan
murata@cs.titech.ac.jp

ABSTRACT

Online social media such as delicious and digg are represented as tripartite networks whose vertices are users, tags, and resources. Detecting communities from such tripartite networks is practically important. Modularity is often used as the criteria for evaluating the goodness of network divisions into communities. Although Newman-Girvan modularity is popular for unipartite networks, it is not suitable for n-partite networks. For bipartite networks, Barber, Guimera, Murata and Suzuki define bipartite modularities. For tripartite networks, Neubauer defines tripartite modularity which extends Murata's bipartite modularity. However, Neubauer's tripartite modularity still uses projections and it will lose information that original tripartite networks have. This paper proposes new tripartite modularity for tripartite networks that do not use projections. Experimental results show that better community structures can be detected by optimizing our tripartite modularity.

Categories and Subject Descriptors

E.1 [Data]: Data Structures—Graphs and networks; G.2.2 [Discrete Mathematics]: Graph Theory—Hypergraphs; D.2.8 [Software Engineering]: Metrics—Performance measures

General Terms

complex network

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 impor-

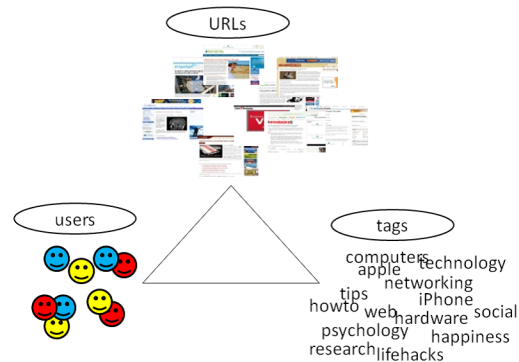tant for finding similar entities and understanding the structure of social media. (Figure 1)

Figure 1: Social Media as a Tripartite Network

As a naive approach for transforming n-partite networks into unipartite networks, projection is often employed for the sake of convenience[16]. 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 [14] is often employed. Optimizing the modularity 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[11] and Suzuki[15]. As an attempt to extend modularity for tripartite networks, Neubauer recently proposes a tripartite

modularity [12] 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 [14] 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_l 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:

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

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:

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

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

$$Q = \sum_l (e_{ll} - a_l^2) \qquad (3)$$

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 modularity maximization algorithm for efficient search for network divisions of high modularity. Newman [13] proposes a spectral algorithm for improving the quality of network division. Blondel et al. [3] attempt community detection from large-scale networks. Danon et al. [6] compare several network division methods. Fortunato et al. [7] clarify resolution limits of modularity-based network division methods.

2.3 Bipartite Modularity

2.3.1 Guimera's Bipartite Modularity

A community is characterized by larger density of intra-community 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:

$$M_B = \sum_{s=1}^{N_M} \left\{ \frac{\sum_{i \neq j \in s} c_{ij}}{\sum_a m_a (m_a - 1)} - \frac{\sum_{i \neq j \in s} t_i t_j}{(\sum_a m_a)^2} \right\} \qquad (4)$$

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:

$$M'_B = \sum_{a=1}^{N'_M} \left\{ \frac{\sum_{i \neq j \in a} c_{ij}}{\sum_s m_s (m_s - 1)} - \frac{\sum_{i \neq j \in a} t_i t_j}{(\sum_s m_s)^2} \right\} \qquad (5)$$

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:

$$A = \begin{bmatrix} 0_{p \times p} & \tilde{A}_{p \times q} \\ (\tilde{A}^T)_{q \times p} & 0_{q \times q} \end{bmatrix}, \qquad (6)$$

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:

$$P = \begin{bmatrix} 0_{p \times p} & \tilde{P}_{p \times q} \\ (\tilde{P}^T)_{q \times p} & 0_{q \times q} \end{bmatrix}, \tag{7}$$

Barber's bipartite modularity is defined as follows:

$$Q_B = \frac{1}{m} \sum_{i=1}^{p} \sum_{j=1}^{q} (\tilde{A}_{ij} - \tilde{P}_{ij}) \delta(g_i, g_{j+p}), \tag{8}$$

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 [11]. 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 particular 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_l^X and V_m^Y are the individual communities that belong to the sets ($V^X = \{V_1^X, ..., V_{L^X}^X\}$, $V^Y = \{V_1^Y, ..., V_{L^Y}^Y\}$). $A(i, j)$ 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_l and V_m are of different types (which means $(V_l \in V^X \wedge V_m \in V^Y) \vee (V_l \in V^Y \wedge V_m \in V^X)$), we can define e_{lm} (the fraction of all edges that connect vertices in V_l to vertices in V_m) and a_l (its row sums) just the same as those in section 2.2.

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

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

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:

$$Q_M = \sum_l Q_{M_l} = \sum_l (e_{lm} - a_l a_m), \quad m = \operatorname*{argmax}_k (e_{lk}) \tag{11}$$

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. [15] 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 [12] 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 $g_X \cup g_Y$ 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.

$$
\begin{aligned}
Q_{3B} = & \frac{1}{3}(Q_M(DU(H), g_D \cup g_U) + Q_M(DT(H), \\
& g_D \cup g_T) + Q_M(UT(H), g_U \cup g_T))
\end{aligned} \tag{12}
$$

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 [12].

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, 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 "being 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 better than random, its modularity is zero.

- 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. $V = \{V^X \cup V^Y \cup V^Z\}$, 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_l^X, V_m^Y, and V_n^Z are the individual communities that belong to the sets ($V^X = \{V_1^X, ..., V_{L^X}^X\}$, $V^Y = \{V_1^Y, ..., V_{L^Y}^Y\}$, $V^Z = \{V_1^Z, ..., V_{L^Z}^Z\}$). 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_l^X, V_m^Y, and V_n^Z are of different types, we can define e_{lmn} (the fraction of all edges that connect vertices in V_l^X, V_m^Y, and V_n^Z) and its sums over three dimensions, such as a_l, a_m, and a_n.

$$e_{lmn} = \frac{1}{M} \sum_{i \in V_l^X} \sum_{j \in V_m^Y} \sum_{k \in V_n^Z} A(i, j, k) \quad (13)$$

$$a_l^X = \sum_m \sum_n e_{lmn}$$
$$= \frac{1}{M} \sum_{i \in V_l^X} \sum_{j \in V^Y} \sum_{k \in V^Z} A(i, j, k) \quad (14)$$

$$a_m^Y = \sum_l \sum_n e_{lmn}$$
$$= \frac{1}{M} \sum_{i \in V^X} \sum_{j \in V_m^Y} \sum_{k \in V^Z} A(i, j, k) \quad (15)$$

$$a_n^Z = \sum_l \sum_m e_{lmn}$$
$$= \frac{1}{M} \sum_{i \in V^X} \sum_{j \in V^Y} \sum_{k \in V_n^Z} A(i, j, k) \quad (16)$$

Suppose $s_X = \sum_l a_l^X$, $s_Y = \sum_m a_m^Y$, and $s_Z = \sum_n a_n^Z$. From the above definitions, it is obvious that $s_X = s_Y = s_Z = \sum_l \sum_m \sum_n e_{lmn} = 1$. As in the case of unipartite networks, if hyperedge connections are made at random, we would have $e_{lmn} = a_l^X a_m^Y a_n^Z$. Therefore, $Q_l^X = \sum_m \sum_n (e_{lmn} - a_l a_m a_n)$, where $m, n = \underset{j,k}{\mathrm{argmax}}(e_{ljk})$, will be zero. On the other hand, if hyperedges from X-vertices are mainly from the vertices in community V_l^X, the value of Q_l^X will be greater than zero. The sum over all communities of V^X is as follows.

$$Q^X = \sum_l Q_l^X$$
$$= \sum_l \sum_m \sum_n (e_{lmn} - a_l a_m a_n) \quad (17)$$
$$m, n = \underset{j,k}{\mathrm{argmax}}(e_{ljk})$$

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 hyper-

edges. 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.

$$
\begin{aligned}
Q^Y &= \sum_m Q_m^Y \\
&= \sum_m \sum_l \sum_n (e_{lmn} - a_l a_m a_n) \qquad (18) \\
&\quad l, n = \operatorname*{argmax}_{i,k}(e_{imk}) \\
Q^Z &= \sum_n Q_l^Z \\
&= \sum_n \sum_l \sum_m (e_{lmn} - a_l a_m a_n) \qquad (19) \\
&\quad l, m = \operatorname*{argmax}_{i,j}(e_{ijn})
\end{aligned}
$$

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

$$
Q = \frac{1}{3}(Q^X + Q^Y + Q^Z) \qquad (20)
$$

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.
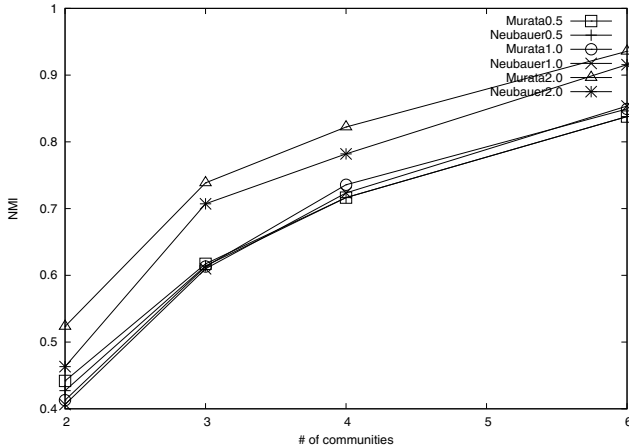
4. EXPERIMENTS

Figure 2: Comparison of Tripartite Modularities

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 ratios of hyperedges per vertices are set to 0.5, 1, and 2, and the numbers of communities in each vertex type are set to 2, 3, 4, and 6. 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. Optimizations are performed in a brute-force, greedy bottom-up fashion.

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],

$$
I(A, B) = \frac{-2\sum_{i=1}^{c_A}\sum_{i=1}^{c_B} N_{ij} log(\frac{N_{ij}N}{N_i.N_{.j}})}{\sum_{i=1}^{c_A} N_i. log(\frac{N_i.}{N}) + \sum_{j=1}^{c_B} N_{.j} log(\frac{N_{.j}}{N})}
$$

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 vertices as $|V^X|(= |V^Y| = |V^Z| = 12)$, the ratio of hyperedges per vertices as $R(=0.5, 1,$ or 2), the number of hyperedges as $M(= |V^X| \cdot R)$, and the number of communities for each vertex type as L ($= L^X = L^Y = L^Z = 2, 3, 4,$ or 6).

2. Assign $|V^X|/L$ vertices to each of the following 'correct' communities: $V_1^X, ... V_L^X, V_1^Y, ... V_L^Y$ and $V_1^Z, ... V_L^Z$.

3. For each (noise-free) hyperedge p ($1 \le p \le M$), vertices $i(\in V_{\lceil (p/M) \cdot L \rceil}^X)$, $j(\in V_{\lceil (p/M) \cdot L \rceil}^Y)$ and $k(\in V_{\lceil (p/M) \cdot L \rceil}^Z)$ are selected randomly from each community and connected with the hyperedge. For a noisy hyperedge, the above vertices are selected from different communities.

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.

Figure 2 shows the results of the optimization of our tripartite modularity and Neubauer's tripartite modularity. X-axis means the number of correct communities and Y-axis means NMI obtained by optimizing tripartite modularities. "Murata0.5" shows the results of optimizing our tripartite

modularity for synthetic tripartite networks whose ratio of hyperedges per vertices is 0.5. Meanings of other legends ("Murata1.0", "Murata2.0", "Neubauer0.5", "Neubauer1.0" and "Neubauer2.0") are explained in the similar way.

Figure 2 shows that the results of optimizing our tripartite modularity are better than those of optimizing Neubauer's tripartite modularity. Optimization of Neubauer's tripartite modularity often results in the detection of too coarse communities since projection often generates too many edges. Other experiments are performed using synthetic tripartite networks with 10% of noisy hyperedges. For noisy tripartite networks also, optimization of our tripartite modularity is better than that of Neubauer's tripartite modularity (Figure 3).
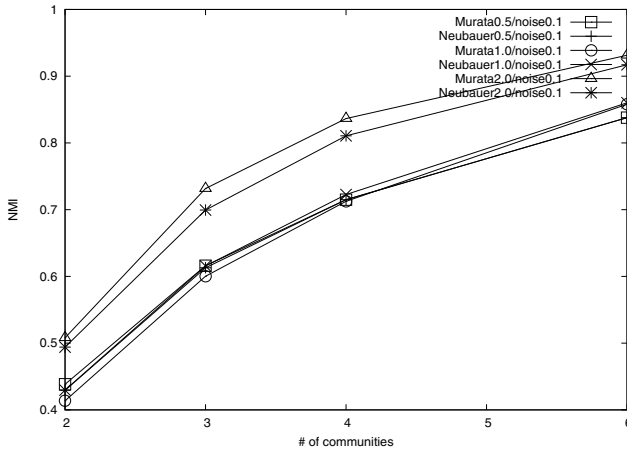

Figure 3: Comparison with Noisy Tripartite Networks

The numbers of communities in tripartite networks are also crucial for the performance. When the number of hyperedges is fixed, connections among corresponding communities are denser for the tripartite networks whose numbers of communities are larger, and Neubauer's tripartite modularity favors the density of the networks. Although the optimization of Neubauer's tripartite modularity are comparable to the optimization of ours for such dense networks, many real heterogeneous 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.

6. REFERENCES

[1] A. Arenas, J. Duch, A. Fernandez, and S. Gomez. Size reduction of complex networks preserving modularity. New Journal of Physics, 9(176):1–15, 2007.

[2] M. J. Barber. Modularity and community detection in bipartite networks. Physical Review E, 76(066102):1–9, 2007.

[3] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of community hierarchies in large networks. Journal of Statistical Mechanics: Theory and Experiment, P10008:1–12, 2008.

[4] C. Cattuto, A. Baldassarri, V. D. Servedio, and V. Loreto. Emergent community structure in social tagging systems. In Proceedings of European Conference on Complex Systems, 2007.

[5] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. Physical Review E, 70(066111):1–6, 2004.

[6] L. Danon, A. Diaz-Guilera, J. Duch, and A. Arenas. Comparing community structure identification. Journal of Statistical Mechanics, P09008:1–10, 2005.

[7] S. Fortunato and M. Barthelemy. Resolution limit in community detection. In Proceedings of the National Academy of Sciences of the United States of America, 104, 1, 36–41, 2007.

[8] R. Ghosh and K. Lerman. Structure of heterogeneous networks. In Proceedings of the 2009 IEEE International Conference on Social Computing (SocialCom-09), 98–105, 2009.

[9] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. Module identification in bipartite and directed networks. Physical Review, 76(036102):1–8, 2007.

[10] E. A. Leicht and M. E. J. Newman. Community structure in directed networks. Physical Review, 100(118703):1–4, 2008.

[11] T. Murata. Detecting communities from bipartite networks based on bipartite modularities. In Proceedings of the 2009 IEEE International Conference on Social Computing (SocialCom-09), 50–57, 2009.

[12] N. Neubauer and K. Obermayer. Towards community detection in k-partite k-uniform hypergraphs. In Proceedings of the NIPS 2009 Workshop on Analyzing Networks and Learning with Graphs, 2009.

[13] M. E. J. Newman. Modularity and community structure in networks. In Proceedings of the National Academy of Sciences of the United States of America, 103, 8577–8582, 2006.

[14] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. Physical Review E, 69(026113):1–15, 2004.

[15] K. Suzuki and K. Wakita. Extracting multi-facet community structure from bipartite networks. In Proceedings of the International Symposium on Social Intelligence and Networking (SIN09), 312–319, 2009.

[16] T. Zhou, J. Ren, M. Medo, and Y.-C. Zhang. Bipartite network projection and personal recommendation. Physical Review E, 76(046115):1–7, 2007.