Detecting Communities from Tripartite Networks

Tsuyoshi Murata
Department of Computer Science, Graduate School of Information Science and Engineering
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. For tripartite networks, Neubauer defines a 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

General Terms

Measurement

Keywords

tripartite network, community, modularity

1. INTRODUCTION

User-resource-tag networks of social tagging systems are the examples of tripartite networks. Detecting communities from such tripartite networks is practically important for finding similar entities and understanding the structure of social media. As a metric for evaluating the goodness of detected communities from unipartite networks, Newman-Girvan modularity is often employed. Since it is not suitable for n-partite networks, Neubauer recently proposes a tripartite modularity [2] based on Murata's bipartite modularity [1]. 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. This paper proposes new tripartite modularity for tripartite networks that does not use projection.

WWW 2010, April 26–30, 2010, Raleigh, North Carolina, USA.
ACM 978-1-60558-799-8/10/04.

2. RELATED WORKS

2.1 Modularity

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. 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:

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

Newman-Girvan modularity is thus defined as follows:

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

2.2 Murata's Bipartite Modularity

N-partite 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 n-partite networks. Murata's bipartite modularity [1] Q_M is defined as follows:

$$Q_M = \sum_l Q_{M_l} = \sum_l (e_{lm} - a_l a_m), \quad m = \underset{k}{\operatorname{argmax}}(e_{lk})$$

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 Neubauer's Tripartite Modularity

Neubauer [2] proposes a tripartite modularity based on Murata's bipartite modularity. 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.

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

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. OUR NEW TRIPARTITE MODULARITY

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. Under the condition that the vertices of V_l^*, V_m^*, and V_n^* are of different types (where $*$ is either X, Y, or Z), we can define e_{lmn} (the fraction of all edges that connect vertices in V_l, V_m, and V_m) and its sums over two dimensions, such as a_l^X, a_m^Y, and a_n^Z.

$$\begin{aligned} e_{lmn} &= \frac{1}{M}\sum_{i \in V_l}\sum_{j \in V_m}\sum_{k \in V_n} A(i,j,k) \\ a_l^X &= \sum_{l \in V_l^X}\sum_{m}\sum_{n} e_{lmn} = \frac{1}{M}\sum_{i \in V_l^X}\sum_{j \in V^Y}\sum_{k \in V^Z} A(i,j,k) \end{aligned}$$

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, $Q_l^X = \sum_m \sum_n (e_{lmn} - a_l^X a_m^Y a_n^Z)$, where $m, n = \underset{i,k}{\operatorname{argmax}}(e_{ljk})$, 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.

$$\begin{aligned} Q^X &= \sum_l Q_l^X = \sum_l \sum_m \sum_n (e_{lmn} - a_l^X a_m^Y a_n^Z) \\ & \quad m, n = \underset{j,k}{\operatorname{argmax}}(e_{ljk}) \end{aligned}$$

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

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

4. EXPERIMENTS

Experiments are performed using synthetic tripartite networks. 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.

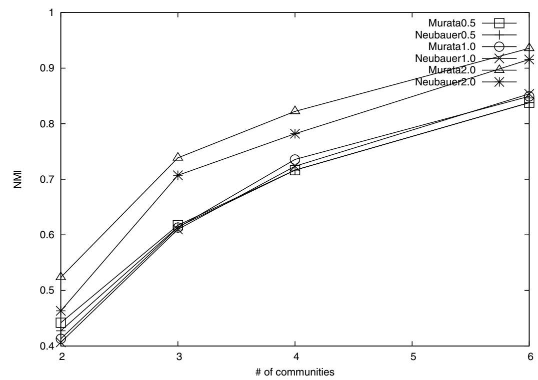

Figure 1: Comparison of Tripartite Modularities

In order to evaluate the goodness of detected communities, normalized mutual information (NMI) is employed. If the detected communities are identical to the correct communities, then NMI takes its maximum value of 1. If the detected communities are totally independent of the correct communities, NMI is 0. "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 are explained in the similar way.

Figure 1 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. For noisy tripartite networks also, optimization of our tripartite modularity is better than that of Neubauer's tripartite modularity.

5. CONCLUSION

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 social tagging systems.

6. REFERENCES

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

[2] 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.