Anchored Maps: Visualization Techniques for Drawing Bipartite Graphs

Kazuo Misue

Department of Computer Science, University of Tsukuba
1-1-1 Tennoudai, Tsukuba, 305-8573 Japan
`misue@cs.tsukuba.ac.jp`

Abstract. A method of drawing anchored maps for bipartite graphs is presented. Suppose that the node set of a bipartite graph is divided into set *A* and set *B*. On an anchored map of the bipartite graph, the nodes in *A*, which are called "anchors," are arranged on the circumference, and the nodes in *B*, which are called "free nodes," are arranged at suitable positions in relation to the adjacent anchors. This article describes aesthetic criteria that are employed according to the purpose of drawing anchored maps and indices which are used to arrange the anchors so that they satisfy the criteria. It also shows an example of taking overviews of networks by using the developed technique.

Keywords: graph drawing, anchored map, bipartite graph, knowledge mining.

1   Introduction

Information with network structures (we call it "network information") can be seen in various fields of the real world. In some fields of research, some of the networks are the research objects such as communication networks and social networks. In recent years, the approach on large-scale networks is active. However, the problem that readability is low in drawing on the networks is not solved still enough. It is an critical problem we should challenge.

The purpose of this research is to develop visualization techniques to give overviews of networks with high readability for mining knowledge. We developed a drawing technique called "anchored map" [5] for the networks which are formalized as bipartite graphs to achieve the purpose. The node set of a bipartite graph is divided into two exclusive sets. Here, we call the nodes in one set "anchors," and the nodes in the other set "free nodes." Our technique arranges anchors on the circumference and arranges free nodes at suitable positions in relation to the adjacent anchors. The anchors are fixed on the circumference as their name suggests and give the effect like the coordinate system to increase readability of the network diagrams. The name "anchored map" was given by the author, but similar representation techniques were also used in some other tools, for example, Visual Who [1], which is a visualization tool for communities and SQWID [3], which is a WWW search tool.

In this article, we describe the technique for drawing anchored maps. First we explain the aesthetic criteria that are employed according to the map's purpose and a method to arrange anchors to fulfill the criteria. Then we propose indices which are

used to arrange the anchors so that they satisfy the criteria and show some evaluation results of their effectiveness. As an example of usage of anchored maps, we show network diagrams which visualize relations between goods and time zones when the goods sold.

2   Anchored Maps of Bipartite Graphs

An anchored map restricts some nodes to certain positions, but leaves other nodes so they may be arranged freely.   The restricted nodes are fixed like anchors, hence the terms "anchors" and "free nodes."

2.1   Bipartite Graphs

In this paper, we focus on networks which can be formalized as bipartite graphs. Suppose that $G = (A \cup B, E)$ is a bipartite graph. A and B are finite sets of vertices, and A and B are disjoint. E is a finite set of edges, and E is a subset of $A \times B$.

2.2   Expression Styles of Bipartite Graphs

Table 1 and Fig. 1 show variations of expression styles of bipartite graphs. Virtual data is used as an example. Suppose that there is a bipartite graph that represents relationships between five communities (C1, ..., C5) and ten members (m01, …, m10). Table 1 shows the matrix representation. Each row corresponds to a community, and each column

Table 1. Matrix Style of a Bipartite Graph

|	C1	C2	C3	C4	C5
m01	1	1			
m02	1			1	1
m03			1		1
m04	1	1		1	
m05	1		1	1	
m06	1	1			
m07			1		1
m08				1	1
m09	1	1		1	
m10		1	1		
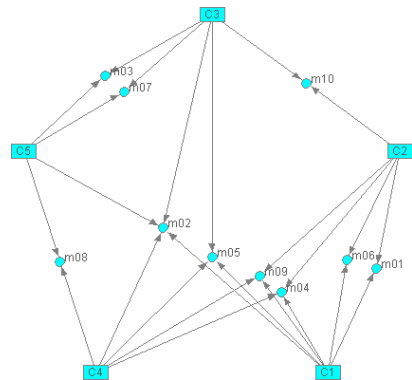
Fig. 1. Anchored-map Style of a Bipartite Graph

corresponds to a member. A "1" in a cell indicates that a member belongs to a community; for example, member m01 belongs to communities C1 and C2. This matrix cannot intuitively show relational structures because it is not visual.

Figure 1 shows the same bipartite graph shown in Table 1 in an anchored-map style. The communities are arranged on the vertices of a pentagon as anchors, and members are arranged at a suitable position to represent their relationship to the

communities as free nodes. The distinction of the nodes is clear, and it is easy to see the relations between members.

2.3   Drawing Procedure of Anchored Maps

An anchored map is laid out in two steps:

(Step 1)   Arrange anchors on the circumference at equal intervals. The size (i.e., radius) of the circumference is decided according to the size of the drawing area (i.e., window), and the order of the anchors on the circumference is decided.

(Step 2)   Fix the anchors, and arrange the free nodes at the positions in which their relationships to the anchors are appropriately expressed.

The length of the circumference doesn't influence the quality of the layout; it influences only the size of the drawing. On the other hand, the order of the anchors has a large influence on the quality of the layout. We do not need to worry about routing of edges since the edges are drawn as straight line segments.

3   Aesthetic Criteria of Anchored Maps

We should provide aesthetic criteria of drawings of networks and develop drawing methods which follow the criteria to enable automatic drawing of networks. We employed the following aesthetic criteria for our anchored map drawing method.

(R1) Nodes are separated mutually more than the lowest distance.

(R2) Nodes connected to each other are laid out as closely as possible (minimize the total length of edges).

(R3) The number of crossings among edges is minimized.

(R4) Anchors connected to common free nodes are laid out as closely as possible.

(R5) Free nodes connected to common anchors are laid out as closely as possible. (Free nodes belonging to the same cluster are laid out as closely as possible; free nodes do not belong to the same cluster are not laid out closely).

The drawing rules (R1), (R2) and (R3) are also employed in many other drawing methods. We also judged them to be important in consideration of the adaptability to the purpose of drawing of anchored maps. The rules (R4) and (R5) are specific to the anchored maps. We show formal description of them in the following part of this section.

3.1   Preparation

Suppose that M is the number of anchors, that is, $M = |A|$. The anchors are arranged on the vertices of a regular M-gon (a polygon with M vertices). The vertices of the M-gon are labeled clockwise from 1 to M. It doesn't matter which vertex is chosen to be 1.

Suppose that $p(a)$ is the position of anchor a and that $A(f)$ is the set of anchors which connect to free node f, that is, $A(f) = \{a \in A \mid (a, f) \in E\}$.

3.2 (R4) Closeness of Anchors Connected to Common Free Nodes

Rule (R4) is to minimize $d(F)$ expressed by

$$d(F) = \sum_{f \in F} d(f) \tag{1}$$

Here, $d(F)$ represents closeness between anchors connected to free node f. Definition of $d(F)$ is described later.

Clockwise Distance. The "clockwise distance" between the i-th vertex and the j-th vertex is the number of vertices we meet when we trace the vertices of the M-gon from the i-th to the j-th clockwise. The clockwise distance l_M is defined as

$$l_M(i, j) = (j - i + M) \bmod M \tag{2}$$

Gap. We introduce a "gap" for every anchor connected to a certain free node. Suppose that f is a free node and $A' = \{a_1, a_2, \cdots, a_k\}$ is a set of anchors connected to f. Moreover, suppose that $a_1, \cdots, a_k$ have been sorted by the numbers of vertices of the M-gon. In other words, $p(a_s) < p(a_t)$ if $s < t$. The gap for each anchor is the clockwise distance between one node and the next node (the next of a_k is a_1). The gap $g(a_s)$ of node a_s is defined as

$$g(a_s) = l_M(p(a_s), p(a_{(s+1) \bmod M})) \tag{3}$$

Closeness of Anchors. When k anchors are connected with a certain free node f, a sequence of k gaps is obtained. Removing the maximum gap from the sequence leaves k-1 smaller gaps. The closeness of anchors is defined by the sum of powers of k-1 gaps and represented by $d(f)$. Suppose that f is a free node and $g_1, g_2, \ldots, g_{k-1}$ is the remainder sequence that excludes the maximum gap from the gap sequence of the anchors connected to f. The closeness of anchors connected to free node f is represented by $d(f)$ and is expressed as

$$d(f) = \sum_{i=1}^{k-1} g_i^q \tag{4}$$

For reason why the parameter q is introduced, please access [4].

Small $d(f)$ indicates that the anchors connected to the same free node should be arranged close to each other.

3.3 (R5) Closeness of Free Nodes Connected to Common Anchors

It is an intuitive and simple rule to make free nodes which connect to the same anchors close to each others. When there are two free node f and g, which connect to the same anchor set, the two free nodes should be arranged close to each other.

However, if there is another free node h, which connects to similar anchor set, we may want to place h a little closer to f and g than the other free nodes which connect completely difference anchors. So we defined similarity between two free nodes by using similarity of sets of anchors connect to the free nodes. The similarity of free nodes f and g is represented by $s(f, g)$ and is expressed as

$$s(f, g) = \frac{|A(f) \cap A(g)|}{|A(f) \cup A(g)|} \tag{5}$$

Rule (R5) can be formalized as to increase correlation between "similarities" and "distance," here similarities are similarities between free nodes and distance is Euclidean distance between the free nodes.

4   How to Decide the Order of Anchors

We employed numeric indices which can be used as replacement of aesthetic criteria (drawing rules) and conducted the experiment that evaluates the effectiveness of the indices. We have introduced an index in [4]. In this paper, we introduce other promised indices and show exhaustive evaluation results.

The distance between adjacent nodes (length of the edges) and the number of edge crossings are essentially influenced by the order of the anchors. However, these features are not deterministically decided from only the anchor order. The spring embedding in Step 2 and the initial positions of the free nodes also influence the distance. However, spring embedding has negligible influence compared with the influence of the anchor order.

4.1   Indices to Decide Anchor Order

The order of the anchors is the most critical for anchored maps. We can evaluate the goodness of a certain order only after the spring embedding has been processed for a certain anchor arrangement, but to do this for all the possible candidate orders would require too much computing time. Thus, we need an alternative index that is computable in a deterministic way and with a low computing cost.

As such the indices we considered to use drawing rules which can be evaluated before arranging anchors. Rules we considered are (R2), (R3) and (R4), and Penalty2, Penalty3 and Penalty4, which described in the followings, are indices correspond to the rules respectively.

(Penalty2) **Total length of edges** when all free nodes are placed at the barycenters of anchors connected to the free nodes (R2).

(Penalty3) **The number of crossings** among edges when all free nodes are placed at the barycenters of anchors connected to the free nodes (R3).

(Penalty4) **The distance along the circumference** between anchors connected to the same free nodes (R4).

We cannot use the drawing rules (R2) and (R3) as they are as indices because both two drawing rules can be evaluated only after the positions of free nodes has been

decided. So, we use the values (edge lengths and number of crossings) which suppose that all free nodes are placed at the barycenters of anchors in definition of Penalty2 and Penalty3. Penalty4 is the same as the Penalty proposed in [4].

4.2   Evaluation of Indices

Table 2 shows the size of the five bipartite graphs used for the evaluation. Table 3 and Table 4 show summary of the evaluation results; Table 3 and Table 4 show the results of Penalty2 and Penalty4 respectively. Both tables show correlations of the penalty with drawing rules and ranks of optimal solution (order) for the indices concerning drawing rules. More precisely, ranks represent the rank of the best anchor order for the indices in all possible orders, that is, 181,440 (= (10-1)! / 2) combinations. There are three values at fields of (R2) and (R3) in the row of G1 in Table 4. It is because that there are three best orders for Penalty4. We had expected that Penalty2 was advantageous to (R2) and Panalty4 was advantageous to (R4). Moreover, we found that Penalty2 was effective for (R3) too.

Table 2. Size of the graphs used for the evaluation

	|A|	|B|	|E|
G1	10	16	65
G2	10	19	84
G3	10	40	182
G4	10	48	251
G5	10	98	473

5   Overview of Network Information

We show an example of overview of network information extracted from an actual database to explain the effectiveness of the anchored maps.

5.1   Database

A sales database is exploited. Table 5 shows a part of the database. One transaction is represented as one record, and one record has purchase date, time, customer name, goods name, goods category and price. We could not get any database of real shops, so we used a sales database managed in our laboratory. It is not a database of a real shop but the transactions in the database are real. We started to use the database in June 2006, and 2652 purchases are recorded for half a year until the end of 2006.

Various relations can be extracted from the database. For example, if we focus on the goods, we can extract relations goods and customers who bought the goods, relations goods and day when the goods was bought, and relations goods and time when the goods was bought. Besides, we can extract relations customers and day when the customers bought, and relations customers and time when the customers bought. Furthermore, we can extract relations customers who bought similar goods, and relations customers who bought at similar time.

Table 3. Evaluation Result of Penalty2

	Correlation with Rules			Ranks of Solutions		
	(r2)	(r3)	(r4)	(r2)	(r3)	(r4)
G1	0.9996	0.9359	0.8949	1	1	4
G2	0.9997	0.9463	0.8873	2	1	756
G3	0.9998	0.9763	0.8912	1	1	657
G4	0.9997	0.9709	0.8416	1	2	62
G5	0.9993	0.9916	0.9025	1	5	1

Table 4. Evaluation Relust of Penalty4

	Correlation with Rules			Ranks of Solutions		
	(r2)	(r3)	(r4)	(r2)	(r3)	(r4)
G1	0.8961	0.8315	1.0000	2,25,244	5,53,566	1
G2	0.8872	0.8240	1.0000	8	27	1
G3	0.8880	0.8592	1.0000	36	20	1
G4	0.8364	0.7969	1.0000	273	680	1
G5	0.9122	0.8954	1.0000	1	5	1

Table 5. Database of Sales (portion)

DATE	TIME	CUSTOMER	GOODS (in Japanese)	CATEGORY	PRICE
2006/6/2	22:32:24	Szyu	ホームラン軒鶏ガラ醤油	convenience food	100
2006/6/5	20:42:47	Masa	午後の紅茶ミルクティー	soft drink	100
2006/6/5	20:37:35	Naga	サッポロ一番塩らーめん	convenience food	100
2006/6/5	20:31:20	KSth	スーパーカップ1.5倍濃コクとんこつ	convenience food	100
2006/6/5	19:39:50	Yuki	昔ながらのソース焼きそば	convenience food	100
2006/6/5	17:59:05	Shuj	カップヌードルカレー	convenience food	100
2006/6/5	16:50:14	Mhro	爽健美茶	tea	100
2006/6/5	16:49:36	Mhro	コアラのマーチチョコレート	sweet	100
2006/6/5	15:09:38	Shuj	Ｂlendy	coffee	100
2006/6/5	14:42:52	Jacq	深い味わい若武者	tea	100
2006/6/5	14:35:13	Snji	コカコーラ	soft drink	100
2006/6/5	13:39:50	Kazu	午後の紅茶ストレートティー	soft drink	100
2006/6/5	12:52:02	Naga	爽健美茶	tea	100
2006/6/5	12:36:17	Snji	爽健美茶	tea	100
2006/6/5	11:21:57	Mhro	午後の紅茶ミルクティー	soft drink	100

5.2   Overview by Using Anchored Maps

Fig. 2(a) shows networks which represent relations between goods and time zone when the goods were bought. Time zones are placed as 24 anchors. Time zone has cyclic order, so they placed in there proper order. From this map, we find at once that
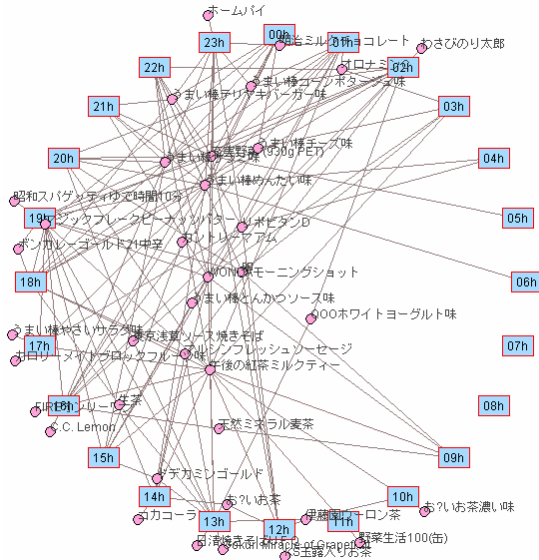
(a) 24 time zones are placed as anchors
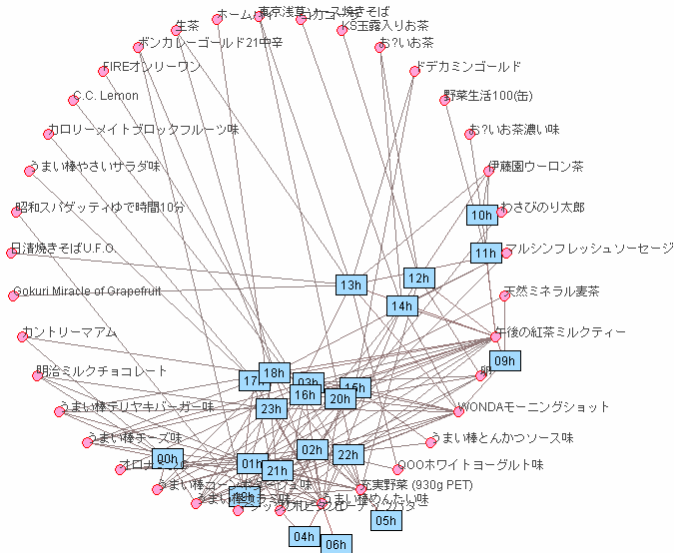
(b) Goods are placed as anchors

Fig. 2. Anchored Maps Showing Relations Goods and Time Zone

the activity of the students is very low in the morning. Furthermore, we find that food such as the curries and spaghettis were bought in specific time zone while the canned coffee and sweets were possibly not depend at specific time zone.

In Fig. 2(b), goods are placed as anchors. Goods have no proper orders, so we use the technique described in Section 4 to decide the order of anchors. We see the following clusters in the map.

(1) Early morning 1: 04h, 05h and 06h
(2) Early morning 2: 07h and 08h (no nodes = no activities)
(3) Morning: 09h, 10h and 11h
(4) Lunch time: 12h, 13h and 14h
(5) Evening ~ night: 15h ~ 03h

It might be possible to find such the clusters without the overview of the networks even by the clustering techniques that already exist. The advantage of overview of network information is that we can see grounds and the background that compose the clusters as some connecting lines between nodes. By using such the maps we can roughly observe the existence of the clusters at first. Then we understand the reason where those clusters exist by touching a virtual physical model of relations.

6   Conclusions

We explained an anchored map drawing technique which we developed to increase readability of network information. We proposed some indices to find good anchor orders effectively and explained effectiveness of the indices by using evaluation results. Moreover, we showed an example of overview of network information extracted from an actual database to illustrate capability of anchored maps at an early stage of observation of network information.

In the example we found some clusters of time zones from the viewpoint of activities of buying. If we would get more rich database we could find more useful relations and background of the relations; relations between goods and area, relations between goods and ages, relations between goods and occupation, and so on.

References

1. Donath, J.S.: Visual Who: Animating the affinities and activities of an electronic community, ACM Multimedia 95. In: Electronic Proceedings (1995)
2. Eades, P.: A Heuristic for Graph Drawing. Congressus Numerantium 42, 149–160 (1984)
3. McCrickard, S.D., Kehoe, C.: Visualizing Search Results using SQWID. In: Proceedings of the sixth International World Wide Web Conference (WWW6) (1997)
4. Misue, K.: Drawing Bipartite Graphs as Anchored Maps. In: Proceedings of Asia-Pacific Symposium on Information Visualization (APVIS 2006), pp. 169–177 (2006)
5. Misue, K., Watanabe, I.: Visualization of Keyword Association for Text Mining. In: IPJS SIG Technical Report, -FI-55-8, Information Processing Society of Japan, (in Japanese) pp. 65–72 (1999)