An experimental comparison of triangulation heuristics on transformed BN2O networks*

Jiří Vomlel
Institute of Information Theory and Automation of the AS CR,
Academy of Sciences of the Czech Republic
http://www.utia.cas.cz/vomlel

Petr Savicky
Institute of Computer Science
Academy of Sciences of the Czech Republic
http://www.cs.cas.cz/savicky

Abstract

In this paper we present results of experimental comparisons of several triangulation heuristics on bipartite graphs. Our motivation for testing heuristics on the family of bipartite graphs is the rank-one decomposition of BN2O networks. A BN2O network is a Bayesian network having the structure of a bipartite graph with all edges directed from the top level toward the bottom level and where all conditional probability tables are noisy-or gates. After applying the rank-one decomposition, which adds an extra level of auxiliary nodes in between the top and bottom levels, and after removing simplicial nodes of the bottom level we get so called BROD graph. This is an undirected bipartite graph. It is desirable for efficiency of the inference to find a triangulation of the BROD graph having the sum of table sizes for all cliques of the triangulated graph as small as possible. From this point of view, the minfill heuristics perform in average better than other tested heuristics (minwidth, h1, and mcs).

1 Introduction

A BN2O network is a Bayesian network having the structure of a bipartite graph with all edges directed from the top level toward the bottom level and where all

*J. Vomlel was supported by grants number 1M0572 and 2C06019 (MŠMT ČR), ICC/08/E010 (Eurocores LogICCC), and 201/09/1891 (GA ČR). P. Savicky was supported by grants number 1M0545 (MŠMT ČR), 1ET100300517 (Information Society), and by Institutional Research Plan AV0Z10300504.

conditional probability tables are noisy-or gates. Let $U = \{u_1, \ldots, u_m\}$ be the nodes of the top level of a BN2O network and $V = \{v_1, \ldots, v_n\}$ be the nodes of the bottom level of this network.

In order to perform efficient inference, we transform these networks using tensor rank-one decomposition [4, 11, 7]. The rank-one decomposition (ROD) graph [8] of a BN2O graph G is the undirected graph constructed from G by

- adding an auxiliary node w_i for each $v_i \in V$,
- replacing each directed edge (u_j, v_i) by an undirected edge $\{u_j, w_i\}$, and
- adding an undirected edge $\{v_i, w_i\}$ for each $v_i \in V$.

The ROD graph is further transformed by triangulation resulting in an undirected triangulated graph. Note that nodes $v_i \in V$ are simplicial in the ROD graph and have degree one. Therefore we can perform optimal triangulation of the ROD graph by optimal triangulation of its subgraph induced by nodes $U \cup W$ [1]. This graph will be called the BROD graph [8]. See Figure 1 for an example of the BROD graph.

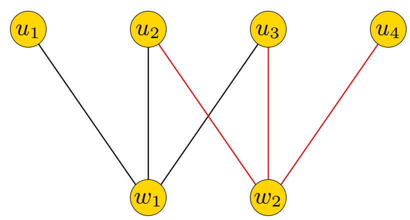

Figure 1: An example of the BROD graph

An important parameter for the inference efficiency is the total table size after triangulation. The table size of a clique C in an undirected graph is $\prod_{v \in C} |X_v|$, where $|X_v|$ is the number of states of a variable X_v corresponding to a node v. If all variables are binary the table size of a clique C is $2^{|C|}$. The total table size of a triangulation is defined as the sum of table sizes for all cliques of the triangulated graph. Therefore, it is desirable to find a triangulation of the BROD graph having the total table size as small as possible. Since this problem is known to be NP-hard and remains NP-hard for bipartite graphs [2], diferent heuristics are often used.

In this paper we perform experimental comparisons of existing heuristic triangulation methods applicable to the BROD graph, which is an undirected bipartite graph. This extends the results already published in [8]. Let us point out that the class of all possible BROD graphs is the same as the class of all bipartite graphs. We talk about BROD graphs, since this corresponds to our motivation.

2 Triangulation heuristics

In Section 3 we will experimentally compare triangulation heuristics minfill [6], minwidth [6], maximum cardinality search [10], and h1 [3]. In order to describe these heuristics, we need notions defined below.

Definition 2.1 Let $G = (V, E)$ be an undirected graph and $U \subseteq V$. The subgraph of G induced by a set of nodes U, denoted $G[U]$, is $G[U] = (U, F)$, where $F = \{\{u, v\} \in E : u, v \in U\}$.

Definition 2.2 Let $F(v) = \{\{v_1, v_2\} : \{v_1, v\} \in E, \{v_2, v\} \in E\}$.

In Table 1 we describe a general template for the considered triangulation heuristics except of minimum cardinality search. The criterion $\phi(u)$ used in step 1 in the template is different for different heuristics and is as follows.

Definition 2.3 *Let v be a node a a graph $G = (V, E)$. Then, let*

1. *$\phi_{minfill}(v)$ be the number of edges added if v is chosen, i.e., $\phi_{minfill}(v) = |F(v) \setminus E|$.*
2. *$\phi_{minwidth}(v)$ be the degree of v, $\phi_{minwidth}(v) = |nb_G(v)|$.*
3. *$\phi_{h1}(v)$ be the size of the largest clique containing $nb_H(v)$, where H is the induced subgraph of $(V, E \cup F(v))$ on the set $V \setminus \{v\}$.*

Table 1: General template for triangulation heuristics using criterion ϕ

For $i = 1, \ldots, |V|$ do:

1. Select a node v of graph G as $v = \arg\min_{u \in V} \phi(u)$, breaking ties arbitrarily.
2. Set $f(v) = i$.
3. Make v a simplicial node in G by adding edges to G, i.e., $G = (V, E \cup F(v))$.
4. Eliminate v from the graph G, i.e. replace G by $G[V \setminus \{v\}]$.

Return f.

Maximum cardinality search has slightly different structure than previously described heuristics. See Table 2.

Table 2: Maximum cardinality search

For all $v \in V$ set weight $w(v) = 0$.
For $i = \|V\|, \ldots, 1$ do:
1. Select an unnumbered node v of graph G maximizing weight w, breaking ties arbitrarily.
2. Set $f(v) = i$.
3. For all unnumbered nodes $u \in nb_G(v)$ set $w(u) = w(u)+1$.
Return f.

3 Experiments

We performed an experimental comparison of the triangulation heuristics on three types of random BN2O graphs. In the first set of experiments, we used 1300 BN2O networks, whose edges were chosen from the uniform distribution on all edges of a complete directed bipartite graph of a given dimension. In the second and third set of experiments we used submodels of the decision theoretic version of Quick Medical Reference (QMR-DT) model using a determinisitic choice of the nodes at the top level and two different types of random choice of the nodes at the bottom level. We will call these submodels QMR thumbnails.

3.1 Randomly generated BN2O networks

First, similarly to [8], we compared the triangulation heuristics on 1300 BN2O networks randomly generated with varying values of the following parameters:

- x, the number of nodes on the top level,
- y, the number of nodes on the bottom level, and
- e, the average number of edges per node on the bottom level.

For each x-y-e type, $x, y = 10, 20, 30, 40, 50$ and $e = 3, 5, 7, 10, 14, 20$ (excluding those with $e \geq x$) we generated randomly ten BN2O graphs by choosing the set of edges from the uniform distribution on the set of all e-tuples of edges from the $x \cdot y$ edges of the complete bipartite graph.

3.2 QMR-DT thumbnails

The decision theoretic version of the Quick Medical Reference [9] (abbreviated QMR-DT) is a large Bayesian network version of the original Quick Medical Reference [5]. There are 570 diseases and 4075 observations in the model. The structure of the model is a directed bipartite graph with edges directed from diseases in the top level to observations in the bottom level. All variables are binary and conditional probability tables of observations given diseases are noisy-or gates. Therefore, QMR-DT represents an example of BN2O model.

Testing triangulation heuristics on the whole QMR-DT is very time consuming and the analysis of this model requires specific algorithms. Our goal is to test the heuristics on smaller graphs. However, we want to test the heuristics on graphs, which contain substructures similar to those, which may appear in real applications. For this purpose, we split the top level of QMR-DT into 10 or 20 disjoint intervals of indices in the order of the nodes, in which the model is presented. This choice implies that similar nodes have higher chance to be chosen to the same subgraph. The exact bounds of the k intervals, where $k = 10$ or $k = 20$, were computed as $[s_{i-1} + 1, s_i]$, where $i = 1, \ldots, k$ and $s_i = \lfloor 570 \cdot i/k \rfloor$.

For each of the k intervals in the top level, denoted X, we used two types of random selection of the set Y of $y = \lceil 4075/k \rceil$ nodes in the bottom level and generated 10 randomly selected sets Y using each of the two methods. Hence, each interval X yields 20 pairs (X, Y) describing a submodel of QMR-DT of the required size. We used the following two types of random selection of Y.

- **Selection by edges.** We choose a random permutation of the edges with the starting point in X from the uniform distribution on such permutations and consider the sequence of the end points of these edges. Then, Y is the set of the first y different nodes in this sequence.

- **Selection by nodes.** We consider the set of end points of the edges, whose starting point is in X. Then, Y is a random subset of these end nodes of size y chosen from the uniform distribution on such subsets.

When $k = 10$, we obtain 200 models, which form the group of thumbnails denoted QMR-DT-57-408. When $k = 20$, we obtain 400 models, which form the group denoted as QMR-DT-29-204.

3.3 Results of experiments

Triangulation heuristics were tested on the BROD graphs G_{BROD}. We used the total table size *tts* of the graph G^h_{BROD} triangulated by a triangulation heuristics h as the criterion for comparisons. We used the *minfill* method as the base method against which we compared all other tested methods. Since randomness is used in the triangulation heuristics we run each heuristics ten times on each model and selected a triangulation with the minimum value of total table size *tts*.

For each tested model we computed the decadic logarithm ratio

$$r(h, minfill) \quad = \quad \log_{10} tts\left(G^h_{BROD}\right) - \log_{10} tts\left(G^{minfill}_{BROD}\right) \;,$$

where h stands for the tested triangulation heuristics.

We used three sets of models for the experiments:

- 1300 randomly generated models *x-y-e* from Section 3.1,
- 200 larger QMR thumbnails *QMR-DT-57-408* from Section 3.2, and
- 400 smaller QMR thumbnails *QMR-DT-29-204* from Section 3.2.

For each of these three groups of models we computed the tts estimate produced by heuristics $h \in \{minfill, minwidth, mcs\}$. For groups x-y-e and QMR-DT-29-204, we additionally computed the triangulation by $h = h1$. The obtained values of tts for $h \neq minfill$ were than compared to the results of $minfill$ for the same group of models. We eliminated the pairs of values of tts for h and $minfill$, which are equal, and performed two-sided Wilcoxon two-sample tests of the null hypothesis that the distribution of $r(h, minfill)$ is symmetric about 0 on the cases, where the two heuristics produced different values. The alternative hypothesis is that the distribution of $r(h, minfill)$ is biased towards negative or positive values. In order to asses, which sign of the typical difference is more likely, we present not only the p-values of the test, but also the values of the statistics W_+ and W_-. If $W_+ > W_-$, then the tested statistics is typically worse than minfill, when $W_+ < W_-$, then it is typically better. The results are summarized in Tables 3, 4, and 5, where nr. obs. means the number of models (observations), for which h and $minfill$ yield different tts.

Table 3: Results of Wilcoxon test for models x-y-e

h	nr. obs.	W_+	W_-	p-val
minwidth	486	78150	40191	8.96e−10
mcs	1266	802011	0	0.00e+00
h1	499	87367	37383	8.88e−15

Table 4: Results of Wilcoxon test for models QMR-DT-57-408

h	nr. obs.	W_+	W_-	p-val
minwidth	193	12946	5775	3.95e−06
mcs	200	20100	0	0.00e+00

The tests revealed that *minfill* performs significantly better than *mcs* on all three sets of models.

Also, *minfill* performed significantly better than *minwidth* on the set of randomly generated models *x-y-e* and on the model set *QMR-DT-57-408*, while on the model set *QMR-DT-29-204* the difference was not significant. On the model set *x-y-e* the advantage of *minfill* over *minwidth* increases with larger value of tts, which was not observed on the other test sets, see Figure 2.

The computations of the *h1* heuristics on *QMR-DT-57-408* took too long, which kept us from the comparisons of *minfill* with *h1* on this model set. On

Table 5: Results of Wilcoxon test for models QMR-DT-29-204

h	nr. obs.	W_+	W_-	p-val
minwidth	313	24517	24624	0.9736
mcs	400	80200	0	0.0000
h1	325	32008	20967	0.0011

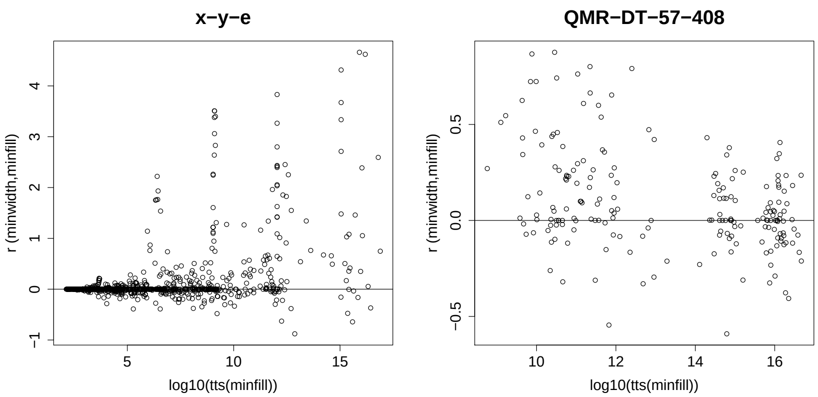

Figure 2: Dependence of $r(minwidth, minfill)$ on decadic logarithm of tts of *minfill* for the set of randomly generated models *x-y-e* and on the model set *QMR-DT-57-408*.

the set of randomly generated models *x-y-e minfill* performed significantly better than *h1* heuristics, while on the model set *QMR-DT-29-204* the difference was not significant.

In Figures 3, 4 and 5 we present histograms of values of $r(h, minfill)$ for *x-y-e*, *QMR-DT-29-204*, and *QMR-57-408* model sets.

4 Conclusions

In this paper we presented results of experimental comparisons of existing heuristic triangulation methods applicable to the BROD graph. The results of experiments reveal that, although no heuristics was dominant on all graphs, in average, the minfill heuristics gave the best results from the tested heuristics.

References

[1] H. L. Bodlaender, A. M. C. A. Koster, and F. Van Den Eijkhof. Preprocessing rules for triangulation of probabilistic networks. *Computational*

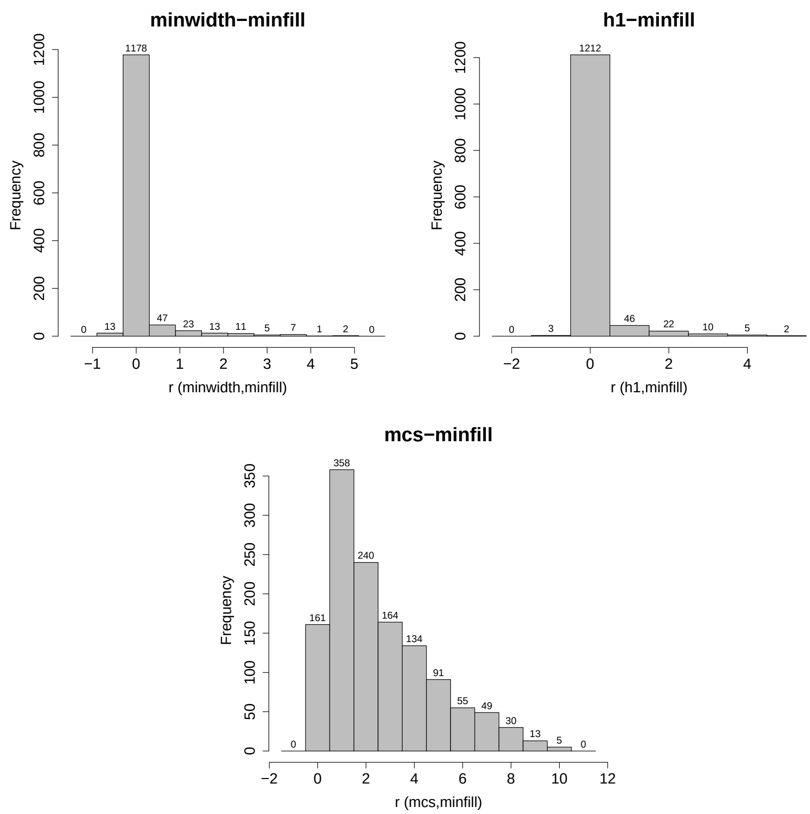

Figure 3: Histograms of values of $r(h, minfill)$ for x-y-e.

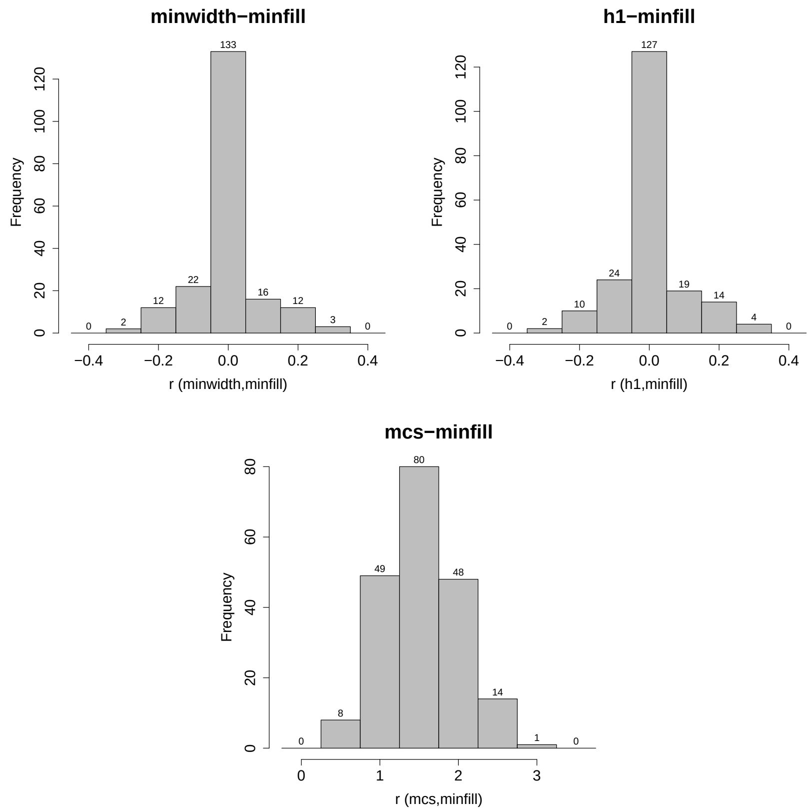

Figure 4: Histograms of values of $r(h, minfill)$ for *QMR-DT-29-204*.

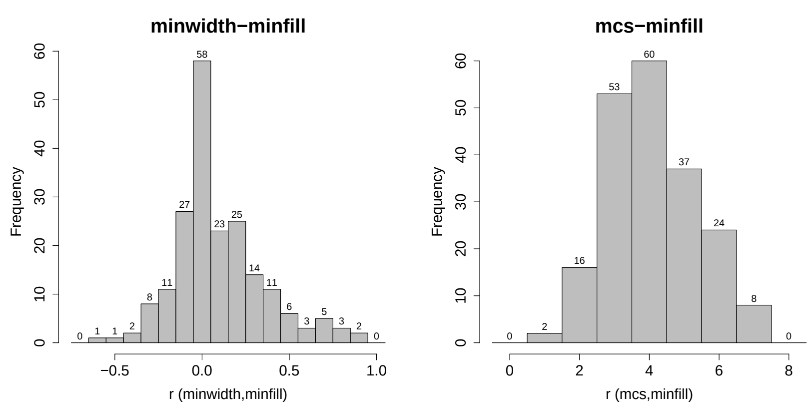

Figure 5: Histograms of values of $r(h, minfill)$ for *QMR-DT-57-408*.

Intelligence, 21(3):286–305, 2005.

[2] Hans L. Bodlaender and Fedor V. Fomin. Tree decompositions with small cost. *Discrete Applied Mathematics*, 145(2):143–154, 2005.

[3] A. Cano and S. Moral. Heuristic algorithms for the triangulation of graphs. In B. Bouchon-Meunier, R. R. Yager, and L. A. Zadeh, editors, *Advances in Intelligent Computing – IPMU '94: Selected Papers*, pages 98–107. Springer, 1994.

[4] F. J. Díez and S. F. Galán. An efficient factorization for the noisy MAX. *International Journal of Intelligent Systems*, 18:165–177, 2003.

[5] R. A. Miller, F. E. Fasarie, and J. D. Myers. Quick medical reference (QMR) for diagnostic assistance. *Medical Computing*, 3:34–48, 1986.

[6] D. J. Rose. A graph-theoretic study of the numerical solution of sparse positive definite systems of linear equations. *Graph Theory and Computing*, pages 183–217, 1972.

[7] P. Savicky and J. Vomlel. Exploiting tensor rank-one decomposition in probabilistic inference. *Kybernetika*, 43(5):747–764, 2007.

[8] P. Savicky and J. Vomlel. Triangulation heuristics for BN2O networks. In C. Sossai and G. Chemello, editors, *Proceedings of the 10th European Conference on Symbolic and Quantitative Approaches to Reasoning with Uncertainty (ECSQARU 2009)*, 2009.

[9] M. Shwe, B. Middleton, D. Heckerman, M. Henrion, E. Horvitz, H. Lehmann, and G. Cooper. Probabilistic diagnosis using a reformulation of the INTERNIST-1/QMR knowledge base. I. The probabilistic model and inference algorithms. *Methods of Information in Medicine*, 30:241–255, 1991.

[10] R. E. Tarjan and M. Yannakakis. Simple linear-time algorithms to test chordality of graphs, test acyclicity of hypergraphs, and selectively reduce acyclic hypergraphs. *SIAM J. Comput.*, 13:566–579, 1984.

[11] J. Vomlel. Exploiting functional dependence in Bayesian network inference. In *Proceedings of the 18th Conference on Uncertainty in AI (UAI)*, pages 528–535. Morgan Kaufmann Publishers, 2002.