Proceedings of the Twenty-Eighth International Florida Artificial Intelligence Research Society Conference

Determining Good Elimination Orderings with Darwinian Networks

Cory J. Butz
University of Regina
Regina, S4S 0A2, Canada
butz@cs.uregina.ca

Jhonatan S. Oliveira
University of Regina
Regina, S4S 0A2, Canada
oliveira@cs.uregina.ca

André E. dos Santos
University of Regina
Regina, S4S 0A2, Canada
dossantos@cs.uregina.ca

Abstract

Darwinian networks (DNs) were recently suggested to simplify reasoning with *Bayesian networks* (BNs). Here we show how DNs can represent four well-known heuristics for determining good elimination orderings in BNs. We propose a new heuristic, called *potential energy* (PE), based on DNs. Our analysis shows that PE compares favourably with these traditional heuristics.

Introduction

Darwinian networks (DNs) (Butz, Oliveira, and dos Santos 2015) were proposed to simplify working with *Bayesian networks* (BNs) (Pearl 1988). DNs are a richer representation allowing them to simplify both modelling and inference.

An important practical consideration in BN inference is determining good *elimination orderings*, denoted σ (Kjærulff 1990). Empirical results show that *min-neighbours* (MN), *min-weight* (MW), *min-fill* (MF), and *weighted-min-fill* (WMF) are four heuristics that perform well in practice (Koller and Friedman 2009). Given a query $P(X|Y)$ posed to a BN $\mathcal{B}$, all variables except XY are recursively eliminated from the moralization $\mathcal{B}_m$ based upon a minimum score $s(v)$.

In this paper, we show how these four heuristics can be represented in DNs. More importantly, we propose a new heuristic, called *potential energy* (PE), based on DNs themselves. Our analysis of PE shows that it can: (i) better score a variable; (ii) better model the multiplication of the probability tables for the chosen variable; (iii) more clearly model the marginalization of the chosen variable; and (iv) maintain a one-to-one correspondence between the remaining variables and probability tables.

Background

Elimination Orderings in Bayesian Networks

Let $U = \{v_1, v_2, \ldots, v_n\}$ be a finite set of variables, each with a finite domain. A singleton set $\{v\}$ may be written as v, $\{v_1, v_2, \ldots, v_n\}$ as $v_1 v_2 \cdots v_n$, and $X \cup Y$ as XY. For disjoint $X, Y \subseteq U$, a *conditional probability table* (CPT) $P(X|Y)$ is a potential over XY that sums to one for each value y of Y.

A *Bayesian network* (BN) (Pearl 1988) is a *directed acyclic graph* (DAG) $\mathcal{B}$ on U together with CPTs $P(v_1|Pa(v_1))$, $P(v_2|Pa(v_2))$,..., $P(v_n|Pa(v_n))$, where the *parents* $Pa(v_i)$ of v_i are those v_j such that $(v_j, v_i) \in \mathcal{B}$. For example, Figure 1 (i) shows a BN, where CPTs $P(a)$, $P(b|a), \ldots, P(f|b, e)$ are not illustrated. We call $\mathcal{B}$ a BN, if no confusion arises.

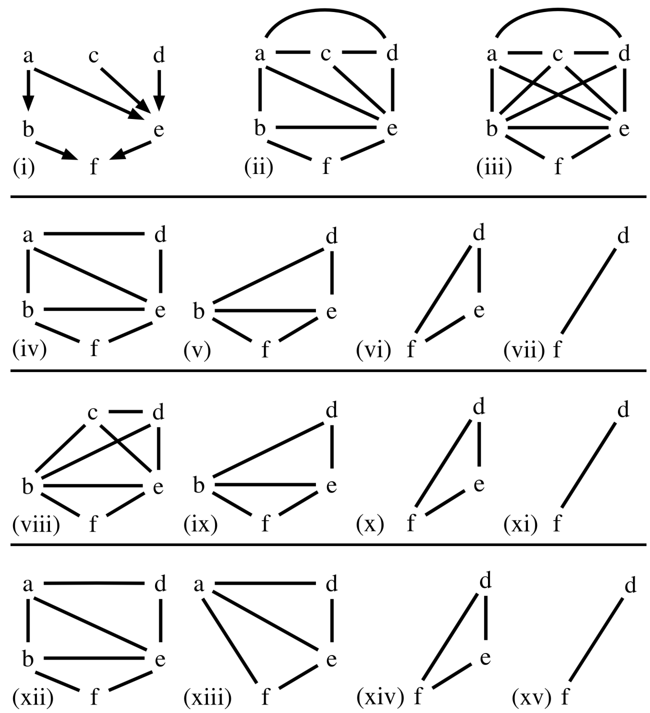

Figure 1: (i) a BN $\mathcal{B}$. (ii) the moralization $\mathcal{B}_m$. (iii) adding edges between a's neighbours in $\mathcal{B}_m$. (iv)-(vii) MN and WMF can determine $\sigma = (c, a, b, e)$. (viii)-(xi) MW can determine $\sigma = (a, c, b, e)$. (xii)-(xv) MF can determine $\sigma = (c, b, a, e)$.

We use the following running example throughout the paper. Assuming query $P(f|d)$ is posed to the BN $\mathcal{B}$ in Figure 1 (i), variables a, b, c, and e must be recursively eliminated from the *moralization* (Pearl 1988) $\mathcal{B}_m$ in Figure 1 (ii). Eliminating a, for instance, involves adding edges between a's neighbours as in (iii) and then removing a as in (viii).

In *min-neighbours* (MN), the score $s(v)$ of variable v is the number of edges involving v.

Example 1. Referring to Figure 1, MN can determine elimination ordering $\sigma = (c, a, b, e)$ by (ii), (iv)-(vii).

In *min-weight* (MW), the score $s(v)$ of a variable v is the product of the domain cardinalities of v's neighbours. The domain cardinality of a is 5, while the rest are binary.

Example 2. Referring to Figure 1, MW can determine $\sigma = (a, c, b, e)$ by (ii), (viii)-(xi).

In *min-fill* (MF), the score $s(v)$ of a variable v is the number of edges that need to be added between v's neighbours due to v's elimination.

Example 3. Referring to Figure 1, MF can determine $\sigma = (c, b, a, e)$ by (ii), (iii), (xii)-(xv).

In *weighted-min-fill* (WMF), the score $s(v)$ of variable v is the sum of the weights of the edges that need to be added between v's neighbours due to v's elimination. The weight of an edge is the product of the domain cardinalities of its constituent vertices.

Example 4. Referring to Figure 1, WMF can determine $\sigma = (c, a, b, e)$ by (ii), (iv)-(vii).

Darwinian Networks

Due to space limitations, we refer the reader to (Butz 2015) for an introduction to DNs. [1]

Every BN $\mathcal{B}$ can be represented as a DN $\mathcal{D}$ (Butz, Oliveira, and dos Santos 2015). More formally, $\mathcal{D} = \{p(v, Pa(v)) \mid P(v|Pa(v))$ is in $\mathcal{B}\}$ is the DN for a given BN $\mathcal{B}$. For instance, the BN $\mathcal{B}$ in Figure 1 (i) is represented as the DN $\mathcal{D}$ in Figure 2 (i).

The query $P(X|Y)$ posed to a BN $\mathcal{B}$ is represented by DN $\mathcal{D}' = \{p(X, Y)\}$. For example, query $P(f|d)$ is represented by DN $\mathcal{D}' = \{p(f, d)\}$ in Figure 2 (viii).

Elimination Orderings in Darwinian Networks

A trait t with a minimum score is eliminated by recursively merging all populations with t, yielding $p(C, D)$, replicating $p(C, D)$ as $p(C, D)$ and $p(C - t, D)$, and letting natural selection remove spent population $p(C, D)$.

Given two DNs $\mathcal{D}$ and $\mathcal{D}'$, recursively eliminate traits appearing in $\mathcal{D}$ but not $\mathcal{D}'$. We will use the same running example for each heuristic, namely, $\mathcal{D}$ is in Figure 2 (i) and $\mathcal{D}'$ is in Figure 2 (viii). By the above, each heuristic seeks to eliminate traits a, b, c, and e from $\mathcal{D}$.

We now present four heuristics for scoring traits in DNs.

Representing Min-Neighbours in DNs

Two traits t and t' in a DN $\mathcal{D}$ are *related*, if they appear together in at least one population in $\mathcal{D}$; otherwise, they are *unrelated*. For example, in Figure 2 (i), trait a is related to traits b, c, d, and e. Traits a and f are unrelated, since they do not appear together in any population.

[1] http://www.darwiniannetworks.com

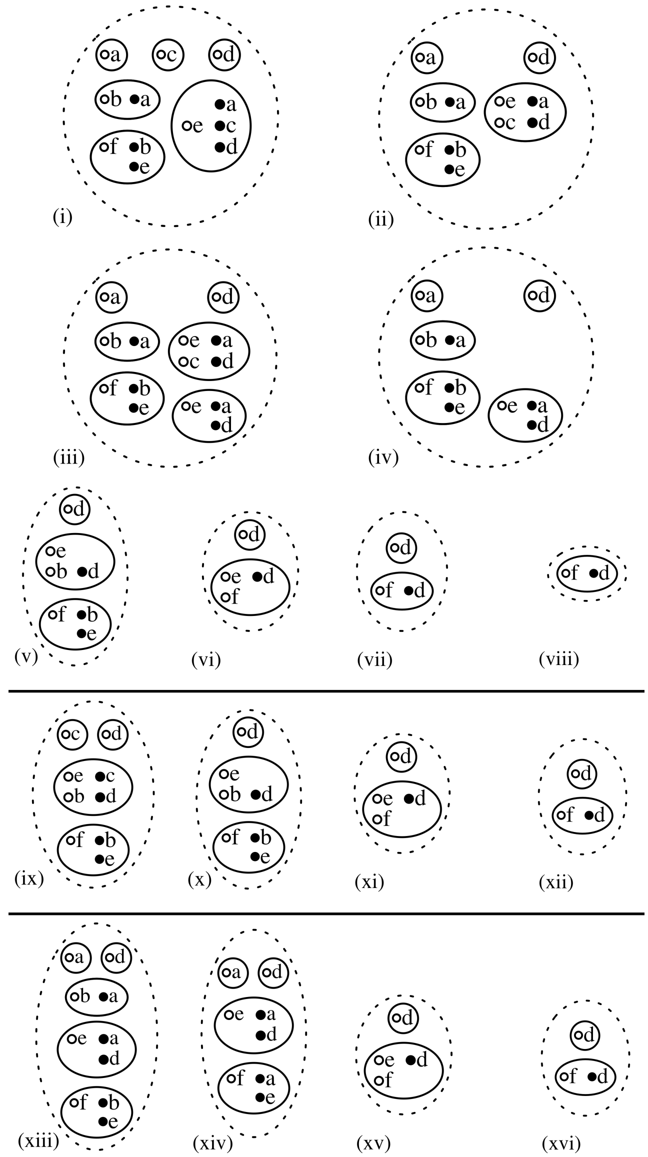

Figure 2: DNs $\mathcal{D}$ in (i) and $\mathcal{D}'$ in (viii). (ii)-(iv) eliminating trait c by merging, replication, and natural selection, respectively. (ii)-(vii) determines $\sigma = (c, a, b, e)$. (ix)-(xii) determines $\sigma = (a, c, b, e)$. (xiii)-(xvi) determines $\sigma = (c, b, a, e)$.

Definition 1. To represent MN, the score $s(t)$ of a trait t is the number of traits related to t.

Example 5. In Figure 2 (i), the score $s(a)$ of trait a is 4, since a is related to traits b, c, d, and e. Similarly, $s(b)$ is 3, $s(c)$ is 3, and $s(e)$ is 5. Eliminating a trait with a minimum score, say c, by merging in (ii), replicating in (iii), and letting natural selection act, gives (iv). Now, $s(a)$ and $s(b)$ both are 3, and $s(e)$ is 4. Eliminating, say a, yields (v). Here, both $s(b)$ and $s(e)$ are 3. Eliminating, say b, gives (vi). Then eliminating e yields (vii). Therefore, $\sigma = (c, a, b, e)$.

Representing Min-Weight in DNs

Given a DN $\mathcal{D}$ representing a BN $\mathcal{B}$, the *energy* of a trait t is the domain cardinality of the variable v to which it corresponds. For example, given that variable b in Example 1 is binary, then the energy of trait b in Example 5 is 2.

Definition 2. To represent MW, the score $s(t)$ of a trait t is the product of the energies of the traits related to t.

Example 6. As the DN $\mathcal{D}$ in Figure 2 (i) represents the BN $\mathcal{B}$ in Figure 1 (i), the energies of traits a, b, c, d, e, and f are 5, 2, 2, 2, 2, and 2, respectively. The traits related to trait a in $\mathcal{D}$ are b, c, d and e. Thus, the score $s(a)$ is 16. Similarly, $s(b)$ and $s(c)$ both are 20, and $s(e)$ is 80. Trait a is eliminated giving (ix). Here, $s(b)$ and $s(e)$ are each 16, while $s(c)$ is 8. Eliminating trait c yields (x). Now, $s(b)$ and $s(c)$ both are 8. Eliminating, say b, gives (xi). Lastly, eliminating trait e yields (xii). Thus, DNs can represent MW determining elimination order $\sigma = (a, c, b, e)$.

Representing Min-Fill in DNs

Definition 3. To represent MF, the score of a trait t is the number of pairs of traits that are related to t, but are unrelated themselves.

Example 7. The traits related to a in $\mathcal{D}$ are b, c, d, and e. Traits b and c are unrelated, as are b and d. Thus, $s(a)$ is 2. Similarly, $s(b)$ is 1, $s(c)$ is 0, and $s(e)$ is 5. Eliminating c gives Figure 2 (xiii). Here, both $s(a)$ and $s(b)$ are 1, and $s(e)$ is 3. Eliminating, say b, yields (xiv). Now, $s(a)$ and $s(e)$ both are 1. Eliminating, say a, gives (xv). Then, e is eliminated leaving (xvi). Therefore, DNs can represent MF determining elimination order $\sigma = (c, b, a, e)$.

Representing Weighted-Min-Fill in DNs

Definition 4. To represent WMF, the score $s(t)$ of a trait t is the product of the energies of the pairs of traits related to t that are unrelated themselves.

Example 8. From Example 7, the two pairs of traits related to a that are unrelated themselves are b and c, and b and d. The product of the energies for b and c is 4, as is that for b and d. Thus, the score $s(a)$ of trait a is 8. Similarly, $s(b)$ is 10, $s(c)$ is 0, and $s(e)$ is 26. Eliminating c gives (iv). Here, $s(a)$ is 4, $s(b)$ is 10, and $s(e)$ is 18. Eliminating a yields (v). Now, $s(b)$ and $s(e)$ both are 4. Eliminating, say b, gives (vi). Finally, e is eliminated yielding (vii). Thus, DNs can represent WMF determining elimination order $\sigma = (c, a, b, e)$.

Equivalence

The *undirected graph* $U(\mathcal{D})$ of a DN $\mathcal{D}$ has variables $T_c(\mathcal{D})$ and edges $\{(v_i, v_j) \mid p(C, D) \in \mathcal{D} \text{ and } v_i, v_j \in CD\}$, namely, $U(\mathcal{D})$ is the moralization $\mathcal{B}_m$ (Butz, Oliveira, and dos Santos 2015). For instance, $U(\mathcal{D})$ of the DN $\mathcal{D}$ in Figure 2 (i) is the undirected graph $\mathcal{B}_m$ shown in Figure 1 (ii).

Lemma 1. MN can be equivalently represented in DNs.

Proof. (Crux) Given $\mathcal{D}$ is the DN for a BN $\mathcal{B}$. Then $U(\mathcal{D})$ is the moralization $\mathcal{B}_m$. By construction, the scoring of traits in $\mathcal{D}$ is precisely the scoring traits in $\mathcal{B}_m$. Also by construction, the undirected graph $U(\mathcal{D})$ of the DN $\mathcal{D}$ obtained by eliminating a trait with a minimum score corresponds exactly to the undirected graph with the corresponding variable eliminated. Finally, the set of traits recursively eliminated is the set of variables recursively eliminated, since $\mathcal{D}' = \{p(X, Y)\}$ is the DN corresponding to the query $P(X|Y)$ posed to $\mathcal{B}$. □

Example 9. The BN $\mathcal{B}$ in Figure 1 (i) is represented as the DN $\mathcal{D}$ in Figure 2 (i), $U(\mathcal{D})$ is the moralization $\mathcal{B}_m$ in Figure 1 (ii), and the given query $P(f|d)$ is represented by the DN $\mathcal{D}' = \{p(f, d)\}$ in Figure 2 (viii). The scoring of variable and traits precisely coincides, namely, $s(a) = 4$, $s(b) = 3$, $s(c) = 3$, and $s(e) = 5$. Eliminating variable c gives Figure 1 (iv), while eliminating trait c gives Figure 2 (iv), of which the undirected graph is Figure 1 (iv). Next, the scoring of variables and traits are the same, namely, both $s(a)$ and $s(b)$ are 3, and $s(e)$ is 4. Eliminating variable a yields Figure 1 (v), while eliminating trait a gives Figure 2 (v), of which the undirected graph is Figure 1 (v). Once again, the scoring is the same, i.e., $s(b)$ and $s(e)$ both are 3. Eliminating variable b gives Figure 1 (vi), while eliminating trait b gives Figure 2 (vi), of which the undirected graph is Figure 1 (vi). Finally, eliminating variable e yields Figure 1 (vii), while eliminating trait e gives Figure 2 (vii), of which the undirected graph is Figure 1 (vii). Therefore, both representations obtain the same elimination ordering $\sigma = (c, a, b, e)$.

Lemma 2. MW, MF, and WMF each can be equivalently represented in DNs.

The proof of Lemma 2 is similar to that of Lemma 1 and will be given in a future paper.

Potential Energy Heuristic

We put forth a new heuristic, called *potential energy*, which is based upon the rich representation of DNs.

Recall that a DN $\mathcal{D}$ can represent a BN $\mathcal{B}$. The *energy of a population* $p(C, D)$ is the domain cardinality of the CPT to which it corresponds. For example, the energy of population $p(b, a)$ representing the CPT $P(b|a)$ is 10, since b is binary and the domain- cardinality of a is 5.

Definition 5. In *potential energy* (PE), the score of a trait t is the sum of the population energies built by recursively merging the populations containing t.

Example 10. Consider using PE to score trait a in the DN $\mathcal{D}$ in Figure 2 (i). Merging $p(a)$ and $p(b, a)$ gives $p(ab)$ with energy 10. Then, merging $p(ab)$ and $p(e, acd)$ gives $p(abe, cd)$ with energy 80. Hence, $s(a)$ is 90.

Note that PE is not necessarily unique, since it depends upon the order in which populations are merged.

Example 11. In Example 10, first merge $p(a)$ with $p(e, acd)$, yielding $p(ae, cd)$ with energy 40. Now, merge $p(ae, cd)$ with $p(b, a)$, giving $p(abe, cd)$ with energy 80. Therefore, $s(a)$ is 120.

Henceforth, populations will be recursively merged always using the two populations with the lowest energies. For example, $s(a)$ will be determined by merging $p(a)$ and

$p(b, a)$ first as in Example 10 rather than $p(a)$ and $p(e, acd)$ as in Example 11.

Example 12. Similar to $s(a)$ being 90 in Example 10, the scores of $s(b)$, $s(c)$, and $s(e)$ are 40, 40, and 160, respectively. Removing, say c, gives Figure 2 (xiii). The scores of $s(a)$, $s(b)$, and $s(e)$ now are 50, 40, and 80, respectively. Removing b yields Figure 2 (xiv). Here, $s(a)$ and $s(e)$ are each 40. Removing, say a, gives Figure 2 (xv), and then e, yields Figure 2 (xvi). Therefore, PE determines elimination ordering $\sigma = (c, b, a, e)$.

Analysis

We show that PE is more refined than MN, MW, MF, and WMF in the following aspects: (i) scoring a variable; (ii) multiplying the probability tables; (iii) marginalizing the variable; and (iv) representing the remaining information.

(i) Scoring a variable. Consider, for example, the multiplications needed to eliminate variable a from the BN in Figure 1 (i):

$$P(a, b, e|c, d) \;=\; P(a) \cdot P(b|a) \cdot P(e|a, c, d). \tag{1}$$

Generally, MF and WMF tend to work better on more problems and, not surprisingly, WMF usually has the most significant gains when there is some significant variability in the sizes of variable domains (Koller and Friedman 2009). This is because MF and WMF estimate the computation to take place in the RHS of (1), for example, by utilizing the RHS itself. PE also considers the RHS of (1), but in more detail than both MF and WMF. PE counts the number of multiplications that can be used to compute the product of all probability tables involving the variable being scored.

Example 13. In Example 10, PE scores a as 90, since computing the product of the RHS of (1) takes 10 multiplications for $p(a)$ and $p(b|a)$, followed by 80 multiplications for $p(a, b)$ and $p(e, acd)$. Thus, a PE score $s(a)$ of 90 means that 90 multiplications can be used to compute (1).

(ii) Multiplying the probability tables. Once a variable with a minimum score is chosen, DNs more accurately depict the multiplication of its probability tables.

Example 14. Suppose c is the first variable to be eliminated in our running example. In the moralization $\mathcal{B}_m$ of Figure 1 (ii), edges are added between all of c's neighbours, yielding $\mathcal{B}_m$ itself. Thus, by using undirected graphs to model multiplication, no change was made in the graphical representation even though the following multiplication takes place:

$$P(c, e|a, d) = P(c) \cdot P(e|a, c, d). \tag{2}$$

In stark contrast, PE explicitly represents this multiplication by merging populations $p(e)$ and $p(e, acd)$ in Figure 2 (i), yielding population $p(ce, ad)$ in Figure 2 (ii).

Example 14 illustrates how merging in DNs is a more descriptive graphical representation of multiplication compared to adding edges in a undirected graph.

(iii) Marginalizing the variable. Once the appropriate probability tables have been multiplied, DNs more accurately depict the subsequent marginalization.

Example 15. Continuing from Example 14, variable c is marginalized in BN inference as follows:

$$P(e|a, d) \;=\; \sum_c P(c, e|a, d). \tag{3}$$

It is not obvious how newly created CPT $P(e|a, d)$ is graphically represented by removing c and its incident edges (a, c), (c, d), and (c, e) from Figure 1 (ii) yielding the undirected graph in Figure 1 (iv). On the other hand, PE clearly articulates this marginalization by replicating population $p(ce, ad)$ in Figure 2 (ii) as itself and population $p(e, ad)$ in Figure 2 (iii), and then letting natural selection remove spent population $p(ce, ad)$ in Figure 2 (iv).

The key point of Example 15 is how replication and natural selection in DNs provide a better graphical description of marginalization in BN inference than the deletion of the variable being marginalized and its incident edges from an undirected graph.

(iv) Representing the remaining information. After a variable with a minimum score is eliminated, PE maintains a one-to-one correspondence between the remaining variables and populations.

Example 16. Continuing from Example 15, c's elimination results in Figure 1 (iv). It is unclear how this undirected graph corresponds to the remaining probability tables:

$$P(a), P(b|a), P(d), P(e|a, d), \text{ and } P(f|b, e). \tag{4}$$

On the contrary, in PE, there is a one-to-one correspondence between the probabilities tables in (4) and populations in Figure 2 (iv):

$$p(a), p(b, a), p(d), p(e, ad), \text{ and } p(f, be). \tag{5}$$

Similar remarks hold after the elimination of variables a, b, and e, in (v), (vi), and (vii) of Figure 2, respectively.

Conclusion

We have shown how four well-known heuristics for determining elimination orderings can be equivalently represented in DNs, a richer representation of BNs. We proposed PE as a novel heuristic based on DNs themselves. Our analysis has shown that PE is a more refined heuristic in four aspects. Future work will include an empirical evaluation.

References

Butz, C. J.; Oliveira, J. S.; and dos Santos, A. E. 2015. Darwinian networks. In *Proceedings of the Twenty-Eighth Canadian Artificial Intelligence Conference (to appear)*.

Butz, C. J. 2015. Introducing darwinian networks. In *Proceedings of the Twenty-Eighth International FLAIRS Conference (to appear)*.

Kjærulff, U. 1990. Triangulation of graphs - algorithms giving small total state space. Technical Report R90-09, Aalborg University, Denmark.

Koller, D., and Friedman, N. 2009. *Probabilistic Graphical Models: Principles and Techniques*. MIT Press.

Pearl, J. 1988. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann.