Final Draft
of the original manuscript:

Kurowicka, D.; Cooke, R.M.; Callies, U.:
Vines inference
In: Brazilian Journal of Probability and Statistics (2007)
Brazilian Statistical Association
2007 Vol.21: 103-120

DOI: -

Vines Inference

D. Kurowicka, R.M. Cooke, U. Callies

April 26, 2006

Abstract

Vines [19, 3] are reviewed. We prove that the product of 1 minus the square of partial correlations on a vine equals the determinant of the correlation matrix. This is used in learning vines. In model learning we are interested in models incorporating maximal (conditional) independence with minimal disturbance. This leads us to search for regular vines whose associated factorization of the determinant is dominant in the sense of majorization. We compare this with the method of independence graphs [21].

1 Introduction

This paper explores the use of vines [19, 3] in graphical modelling. Sampling algorithms for conditional rank correlation vines are presented in section 3. We prove that when partial correlations are assigned to the nodes of a regular vine, then the product of 1 minus the square partial correlations equals the determinant of the correlation matrix. Hence a regular vine can be viewed as a factorization of the determinant. In model learning we are interested in models that incorporate maximal (conditional) independence with minimal disturbance. This leads us to search for regular vines whose associated factorization is dominant in the sense of majorization. Setting small magnitude partial correlations equal to zero thus minimally perturbs the determinant. We compare this with the method of independence graphs [21]. An example from European weather stations illustrates that the vine-based approach can exhibit good performance.

2 Vines - Basic definitions

Graphical models called *vines* were introduced in [19, 3]. A vine $\mathcal{V}(n)$ on n variables is a nested set of trees $\mathcal{V}(n) = (T_1, ..., T_{n-1})$ where the edges of tree j are the nodes of tree $j+1$, $j = 1, ..., n-2$ and each tree has the maximum number of edges. A *regular vine* on n variables is a vine in which two edges in tree j are joined by an edge in tree $j+1$ only if these edges share a common node, $j = 1, ..., n-2$. The formal definitions follow.

Definition 2.1 (Tree) *$T = (N, E)$ is a* **tree** *with nodes N and edges E if E is a subset of unordered pairs of N with no cycle. Moreover, for any $a, b \in N$ there exists a sequence $c_1, ..., c_k$ of elements of N such that*

$$\{a, c_1\} \in E, \{c_1, c_2\} \in E, \ldots, \{c_k, b\} \in E.$$

Definition 2.2 (Regular vine) *V is a* **regular vine** *on n elements if*

1. $\mathcal{V}(n) = (T_1, \ldots, T_{n-1})$,
2. *T_1 is a tree with nodes $N_1 = \{1, \ldots, n\}$, and edges E_1;*
   for $i = 2, \ldots, n-1$ T_i is a tree with nodes $N_i = E_{i-1}$.
3. **(proximity)** *for $i = 2, \ldots, n-1$, $\{a, b\} \in E_i, \#a \triangle b = 2$ where $\triangle$ denotes the symmetric difference and $\#A$ denotes the number of elements in set A.*

A regular vine is called a *canonical vine* if each tree T_i has a unique node of degree[1] $n-i$, hence has maximum degree. A regular vine is called a *D-vine* if all nodes in T_1 have degree not higher than 2 (see Figure 1). There are $n(n-1)/2$ edges in a regular vine on n variables. An edge in tree T_j is an unordered pair of nodes of T_j, or equivalently, an unordered pair of edges of T_{j-1}. By definition, the *order* of an edge in tree T_j is $j-1$, $j=1,...,n-1$. Equivalently the order of a node in T_j is j.

A regular vine is just a way of identifying a set of conditional bivariate constraints. The conditional bivariate constraint associated with each edge are determined as follows: the variables reachable from a given edge via the membership relation are called the *constraint set* of that edge. When two edges are joined by an edge of the next tree, the intersection of the respective constraint sets are the *conditioning variables*, and the symmetric differences of the constraint sets are the *conditioned variables*. The order of an edge, or equivalently of a node, is the cardinality of its conditioning set.

Definition 2.3 (Conditioning, conditioned and constraint sets)

1. *For $j \in E_i, i \leq n-1$ the subset $U_j(k)$ of $E_{i-k} = N_{i-k+1}$, defined by $U_j(k) = \{e \mid \exists\ e_{i-(k-1)} \in e_{i-(k-2)} \in \ldots \in j,\ \ e \in e_{i-(k-1)}\}$, is called the* **k-fold union**
$U_j^* = U_j(i)$ is the **complete union** *of j, that is, the subset of $\{1,\ldots,n\}$ consisting of $m-$descendants of j.*
If $a \in N_1$ then $U_a^* = \emptyset$.
$U_j(1) = \{j_1, j_2\} = j$.
By definition we write $U_j(0) = \{j\}$.

2. *The* **constraint set** *associated with $e \in E_i$ is U_e^*.*

3. *For $i = 1,\ldots,n-1, e \in E_i$, $e = \{j,k\}$, the* **conditioning set** *associated with e is*

$$D_e = U_j^* \cap U_k^*$$

and the **conditioned set** *associated with e is*

$$\{C_{e,j}, C_{e,k}\} = U_j^* \triangle U_k^* = \{U_j^* \setminus D_e,\ U_k^* \setminus D_e\}.$$

The order of node e is $\#D_e$.

Note that for $e \in E_1$, the conditioning set is empty.
For $e \in E_i, i \leq n-1, e = \{j,k\}$ we have $U_e^* = U_j^* \cup U_k^*$.

We use the D-vine in Figure 1 to illustrate Definition 2.3.
We get

$$\begin{aligned}
T_1 = (N_1, E_1), \qquad & N_1 = \{1,2,3,4\}, \\
& E_1 = \{(1,2);(2,3);(3,4)\}; \\
T_2 = (N_2, E_2), \qquad & N_2 = E_1, \\
& E_2 = \{((1,2),(2,3))\ ;\ ((2,3),(3,4))\}; \\
T_3 = (N_3, E_3) \qquad & N_3 = E_2, \\
& E_3 = \{(((1,2),(2,3)),((2,3),(3,4)))\};
\end{aligned}$$

The complete union of $j = \{1,2\}$ is $U_j^* = \{1,2\}$, and for $k = \{2,3\}$, $U_k^* = \{2,3\}$. Hence the conditioning set of the edge $e = \{\{1,2\},\{2,3\}\}$ in T_2 is $D_e = U_j^* \cap U_k^* = \{1,2\} \cap \{2,3\} = \{2\}$. The

[1]The degree of the node $a \in N$ is the number of edges attached to a.

conditioned set consists of $C_{e,j} = U_j^* \setminus D_e = \{1,2\} \setminus \{2\} = \{1\}$ and $C_{e,k} = U_k^* \setminus D_e = \{2,3\} \setminus \{2\} = \{3\}$. Hence the conditioned set of $e = ((1,2),(2,3))$ is $\{1,3\}$ and the conditioning set $\{2\}$. The conditioning and conditioned sets of the single edge in T_3 in the D-vine in Figure 1 are $\{2,3\}$ and $\{1,4\}$, respectively.
The following proposition is proved in [3, 13]:

Proposition 2.1 *Let $\mathcal{V}(n) = (T_1, ...T_{n-1})$ be a regular vine, then*

1. *the number of edges is $\frac{n(n-1)}{2}$,*
2. *each conditioned set is a doubleton, each pair of variables occurs exactly once as a conditioned set,*
3. *if two edges have the same conditioning set, then they are the same edge,*
4. *if $e \in E_i, i = 1, \ldots n-1$, then $\#D_e = i - 1$.*

The results in this paper require in addition:

Proposition 2.2 *For any node e of order $k > 0$ in a regular vine, if variable i is a member of the conditioned set of e, then i is a member of the conditioned set of exactly one of the m-children of e, and the conditioning set of an m-child of e is a subset of the conditioning set of e.*

Proof. If the conditioning set of an m-child of e is vacuous, the proposition is trivially true, we therefore assume $k > 1$. Let $e = \{a, b\}$, where a, b are nodes of order $k-1$. By regularity we may write $a = \{a1, d\}, b = \{b1, d\}$ where $a1, b1, d$ are nodes of order $k-2$. Let C_x denote the constraint set of node x, we write $C_e = C_a \cup C_b$. By assumption, $i \in C_a \triangle C_b$. Suppose $i \in C_a$, then $i \notin C_b$. $C_a = C_{a1} \cup C_d$, and since $C_d \subseteq C_b$ and $i \notin C_b$, we have $i \notin C_d$. It follows that $i \in C_{a1} \triangle C_d$; that is, i is in the conditioned set of a. Since the conditioning set of a is $C_{a1} \cap C_d \subseteq C_b$, we have $C_{a1} \cap C_d \subseteq C_a \cap C_b$; that is, the conditioning set of a is a subset of the conditioning set of e.□

When vines are used to specify a distribution and design a sampling procedure, the bivariate constraints must be associated with conditional bivariate distributions. This is the subject of the next section. When learning a model from data, it is convenient to think of these bivariate constraints in terms of second moment structures.

3 Sampling vines

We assume that variables $X_1, X_2, ..., X_n$ are uniform on $(0,1)$. Each edge in a regular vine may be associated with a conditional copula, that is, a conditional bivariate distribution with uniform margins (for j=1 the conditions are vacuous). It is convenient to specify the conditional bivariate copulae by first assigning a constant conditional rank correlation to each edge of the vine. For $i = 1, \ldots, n-1$, with $e \in E_i$ and $\{j, k\}$ the conditioned variables of e, D_e the conditioning variables of e, we associate

$$r_{j,k|D_e}.$$

The resulting structure is called a *conditional rank correlation vine*. Given a conditional rank correlation vine, we choose a class of copulae indexed by correlation coefficients in the interval $[-1, 1]$ and select the copulae with correlation corresponding to the conditional rank correlation assigned to the edge of the vine. A joint distribution satisfying the vine-copula specification can be constructed and sampled on the fly, and will preserve maximum entropy properties of the conditional bivariate distributions [19, 2].

The conditional rank correlation vine plus copula determines the whole joint distribution. When the choice of copula is unrestricted, there are two strategies for sampling such a distribution, which we term the *cumulative* and *density* approaches [2].

[2] If we choose the''joint normal copula' then a simpler sampling strategy exists

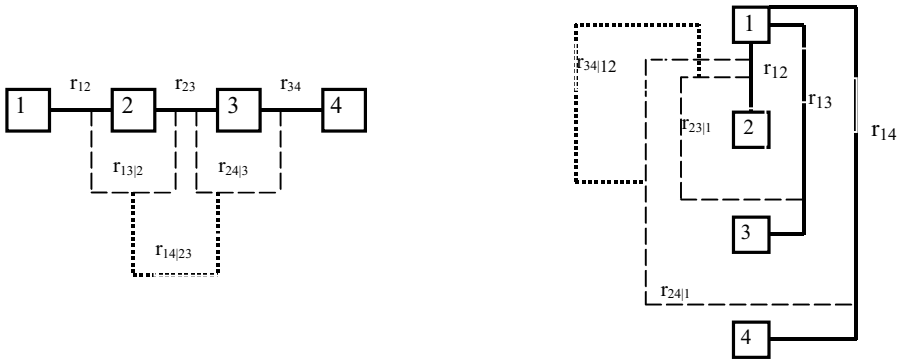

Figure 1: D-vine (left) and canonical vine (right) on 4 variables with (conditional) rank correlations assigned to the edges.

3.1 Cumulative approach

We first illustrate the cumulative approach with the distribution specified by the D-vine in Figure 1: Sample four independent variables distributed uniformly on interval [0,1], U_1, U_2, U_3, U_4 and calculate values of correlated variables X_1, X_2, X_3, X_4 as follows:

1. $x_1 = u_1$,
2. $x_2 = F^{-1}_{r_{12};x_1}(u_2)$,
3. $x_3 = F^{-1}_{r_{23};x_2}\left(F^{-1}_{r_{13|2};F_{r_{12};x_2}(x_1)}(u_3)\right)$,
4. $x_4 = F^{-1}_{r_{34};x_3}\left(F^{-1}_{r_{24|3};F_{r_{23};x_3}(x_2)}\left(F^{-1}_{r_{14|23};F_{r_{13|2};F_{r_{23};x_2}(x_3)}(F_{r_{12};x_2}(x_1))}(u_4)\right)\right)$

where $F_{r_{ij|k};X_i}(X_j)$ denotes the cumulative distribution function for X_j, applied to X_j, given X_i under the conditional copula with correlation $r_{ij|k}$. Notice that the D-vine sampling procedure uses conditional and inverse conditional distribution functions. A more general form of the above procedure simply refers to conditional cumulative distribution functions:

$$\begin{aligned} x_1 &= u_1, \\ x_2 &= F^{-1}_{2|1:x_1}(u_2), \\ x_3 &= F^{-1}_{3|2:x_2}\left(F^{-1}_{3|12:F_{1|2}(x_1)}(u_3)\right), \\ x_4 &= F^{-1}_{4|3:x_3}\left(F^{-1}_{4|23:F_{2|3}(x_2)}\left(F^{-1}_{4|123:F_{1|23}(x_1)}(u_4)\right)\right). \end{aligned} \tag{1}$$

Figure 2 depicts the sampling of X_4 in the D-vine in Figure 1. In the sampling procedure the diagonal band copula [5] was used. We assume that we have already sampled X_1, X_2 and X_3. Hence the values x_3, $F_{2|3}$ and $F_{1|23}$ are known. Conditionalizing the copulae with correlations $r_{23}, r_{24|3}$ and $r_{14|23}$ on x_3, $F_{2|3}$ and $F_{1|23}$, respectively one can find conditional cumulative distributions $F_{4|3}, F_{4|23}$ and $F_{4|123}$. Inverting value of u_4 with $F_{4|123}$ we obtain value of $F_{4|23}$, then this value is inverted using $F_{4|23}$ to get the value of $F_{4|3}$. Finally x_4 is calculated by inverting $F_{4|3}$.

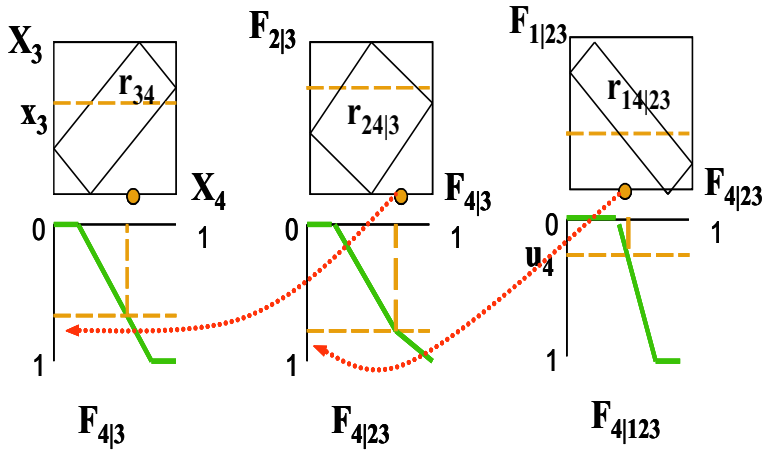

Figure 2: Sampling value of X_4 in the D-vine on four variables.

A *general cumulative sampling algorithm* for an n-dimensional distribution represented graphically by the D- vine on n variables is given as follows

$$\begin{aligned}
x_1 &= u_1, \\
x_2 &= F^{-1}_{2|1:x_1}(u_2), \\
x_3 &= F^{-1}_{3|2:x_2}\left(F^{-1}_{3|12:F_{1|2}(x_1)}(u_3)\right), \\
x_4 &= F^{-1}_{4|3:x_3}\left(F^{-1}_{4|23:F_{2|3}(x_2)}\left(F^{-1}_{4|123:F_{1|23}(x_1)}(u_4)\right)\right), \\
x_5 &= F^{-1}_{5|4:x_4}\left(F^{-1}_{5|34:F_{3|4}(x_3)}\left(F^{-1}_{5|234:F_{2|34}(x_2)}\left(F^{-1}_{5|1234:F_{1|234}(x_1)}(u_5)\right)\right)\right), \\
&\dots \\
x_n &= F^{-1}_{n|n-1:x_{n-1}}\left(F^{-1}_{n|n-2,n-1:F_{n-2|n-1}(x_{n-2})}\left(F^{-1}_{n|n-3,n-2,n-1:F_{n-3|n-2,n-1}(x_{n-3})}(\dots\right.\right. \\
&\quad \left.\left.\left(F^{-1}_{n|1\dots n-1:F_{1|2\dots n-1}(x_1)}(u_n)\right)\dots\right)\right);
\end{aligned} \tag{2}$$

When the bivariate distributions are indexed by conditional rank correlations, the correlations to be specified are

$$\begin{array}{ccccccc}
r_{12}, & r_{13|2}, & r_{14|23}, & r_{15|234}, & \dots & r_{1,n-1|2\dots n-2}, & r_{1,n|2\dots n-1}, \\
 & r_{23}, & r_{24|3}, & r_{25|34}, & \dots & r_{2,n-1|3\dots n-2}, & r_{2,n|3\dots n-1}, \\
 & & r_{34}, & r_{35|4}, & \dots & r_{3,n-1|4\dots n-2}, & r_{3,n|4\dots n-1}, \\
 & & & & \dots & & \\
 & & & & & r_{n-2,n-1}, & r_{n-2,n|n-1}, \\
 & & & & & & r_{n-1,n},
\end{array} \tag{3}$$

Notice that the conditional rank correlations can be chosen arbitrarily in the interval $[-1, 1]$; they need not be positive definite or satisfy any further algebraic constraint.

3.2 Density approach sampling

When the vine-copula distribution is given as a density, the density approach to sampling may be used. If E_m is the edge set for tree T_m, and for $e \in E_m$ with conditioning set D_e, let $c_{ij|D_e}$ be the copula density associated with e, then the density for a distribution specified by the assignment of copulae to the edges of $\mathcal{V}(n)$ is given by [2]:

$$\prod_{m=1}^{n-1}\prod_{e\in E_m} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) f_1(x_1)\dots f_n(x_n) = \prod_{m=1}^{n-1}\prod_{e\in E_m} c_{ij|D_e}(F_{i|D_e}(x_i), F_{j|D_e}(x_j)) \tag{4}$$

where, by uniformity, the density $f_i(x_i) = 1$.

This expression may be used to sample the vine distribution; namely, draw a large number of samples $(x_1, \ldots, x_n)$ uniformly, and then resample these with probability proportional to (4). This is less efficient than the general sampling algorithm given previously; however it may be more convenient for conditionalization.

4 Model Learning

We consider a situation in which we have data from a multivariate distribution on n variables. We wish to find an appropriate model structure based on this data. We assume that the data is not sufficiently rich to allow estimates of conditional bivariate distributions, rather our inferences will be based on the empirical univariate distributions and the empirical covariance matrix. We assume that the univariate distributions are empirical samples of the uniform $[0, 1]$ distribution.

We shall approach model learning as a problem of inferring a model with certain desirable properties from data. We first review the theory of independence graphs in this regard, as this provides a point of departure for the problem of vine inference. We shall associate nodes in a regular vine with *partial correlations*; partial correlation $\rho_{ij;K}$ may be associated with a node in a regular vine if K is the conditioning set, and $\{i, j\}$ the conditioned set of the node. The result of such an association is a *partial correlation vine.*

Partial correlation can be defined in terms of partial regression coefficients. Let us consider variables X_i with zero mean and standard deviations σ_i, $i = 1, \ldots, n$. Let the numbers $b_{12;3,\ldots,n}, \ldots,$ $b_{1n;2,\ldots,n-1}$ minimize

$$E\left((X_1 - b_{12;3,\ldots,n}X_2 - \ldots - b_{1n;2,\ldots,n-1}X_n)^2\right).$$

Definition 4.1 (Partial correlation)

$$\rho_{12;3,\ldots,n} \quad = \quad sgn(b_{12;3,\ldots,n})\left(b_{12;3,\ldots,n}b_{21;3,\ldots,n}\right)^{\frac{1}{2}}, \ \textit{etc.}$$

Equivalently we could define the partial correlation as

$$\rho_{12;3,\ldots,n} \quad = \quad \frac{C_{12}}{\sqrt{C_{11}C_{22}}},$$

where $C_{i,j}$ denotes the $(i, j)th$ cofactor of the correlation matrix.
The partial correlation $\rho_{12;3,\ldots,n}$ can be interpreted as the correlation between the orthogonal projections of X_1 and X_2 on the plane orthogonal to the space spanned by $X_3, \ldots, X_n$.

Partial correlations can be computed from correlations with the following recursive formula [22]:

$$\rho_{12;3,\ldots,n} \quad = \quad \frac{\rho_{12;3,\ldots,n-1} - \rho_{1n;3,\ldots,n-1} \cdot \rho_{2n;3,\ldots,n-1}}{\sqrt{1 - \rho^2_{1n;3,\ldots,n-1}}\sqrt{1 - \rho^2_{2n;3,\ldots,n-1}}}. \tag{5}$$

The main fact about partial correlation specification on vines is given in [3]:

Theorem 4.1 *For any regular vine on n variables, any assignment of values in the interval* $(-1, 1)$ *to the partial correlations associated with the edges of the vine defines a unique correlation matrix and every correlation matrix with rank n arises in this way.*

In other words, a partial correlation vine fully characterizes the correlation structure of the joint distribution and the values of the partial correlations are algebraically independent. Unlike the values in a correlation matrix, the partial correlations in a regular vine need not satisfy an algebraic constraint like positive definiteness.

It is well known that for the joint normal distribution, partial and conditional correlation coincide. This is not true in general. The degree to which constant conditional rank correlations approximate partial correlations depends strongly on the choice of copula [14].

The approach in this section is inspired by that of [1, 12] and especially [21], which we briefly review.

4.1 Inference with Independence graphs

In [21] the problem of model inference is cast as a problem of identifying conditional independence. The joint distribution of variables $1, ..., n$ is assumed to be joint normal. The approach is sketched as follows:

1. Estimate the variance V by the sample variance matrix S.
2. Compute its inverse S^{-1}, re-scale this inverse so that the diagonal entries are 1; the off-diagonal cell $\{i, j\}$ contains the negative of the partial correlation of i, j with respect to all remaining variables K.
3. Set any sufficiently small partial correlations in S^{-1} equal to zero; call the resulting matrix P^*[3].
4. Find a positive definite matrix P 'as close as possible' to P^*; the zero's of P correspond correspond to pairs of variables which are modelled as conditionally independent given all other variables.

The procedure for finding the positive definite matrix P uses iterative proportional fitting, and relies heavily on the properties of the joint normal distribution. P and P^* have the same zeros. Goodness of fit tests are available to determine whether P^{-1} is 'close enough' to the original sample variance S.

The resulting structure is expressed as an 'independence graph'. Starting with the saturated undirected graph over all variables, edges between variables are removed if the partial correlation of these variables, given the remaining variables, is zero. The remaining edges connect variables having an interaction in the inferred model.

4.2 Vine Inference

In analogy with the independence graph method, we look for a partial correlation vine which contains as many zero's as possible while remaining as close as possible to the empirical distribution. This approach utilizes properties of vines and the partial correlations vines that are developed below. A strategy for vine inference is then suggested.

Partial correlation vines The following theorem uses the multiple correlation [11]:

Definition 4.2 (Multiple correlation) *The multiple correlation* $R_{1\{2,...,n\}}$ *of variable* 1 *with respect to* $2, ..., n$ *is*

$$R_{1\{2,...,n\}} = \frac{D}{C_{11}}$$

where D *is the determinant of the correlation matrix. It is the correlation between* 1 *and the best linear predictor of* 1 *based on* $2, ..., n$.

[3]The process of removing links corresponding to small partial correlations can be carried out sequentially. After each removal the correlation matrix and partial correlations are updated. The decision to remove a link is based on the (a) 'edge exclusion-' and (b) 'edge inclusion-deviance'. These are entropy based measures that reflect (a) the additional disturbance of the original distribution caused by removing an additional edge and (b) the reduction in disturbance achieved by restoring previously removed edge. The process stops when the minimum edge exclusion deviance is much bigger then the maximum edge inclusion deviance.

$R_{1\{2,...,n\}}$ is non negative and satisfies:

$$1 - R^2_{1\{2,...,n\}} = (1-\rho^2_{1,n})(1-\rho^2_{1,n-1;n})(1-\rho^2_{1,n-2;n-1,n})...(1-\rho^2_{1,2;3...n}). \quad (6)$$

It follows from ([11] p. 324), that $R_{1\{2,...,n\}}$ is invariant under permutation of $\{2,...,n\}$ and

$$D = \left(1 - R^2_{1\{2,...,n\}}\right)\left(1 - R^2_{2\{3,...,n\}}\right)...\left(1 - R^2_{n-1\{n\}}\right); \quad (7)$$

Of course $R_{n-1\{n\}} = \rho_{n-1,n}$.

Theorem 4.2 *Let D be the determinant of the correlation matrix of variables $1,...,n$; $D > 0$. For any partial correlation vine;*

$$D = \prod_{\{i,j\}} (1 - \rho^2_{i,j;K(ij)}) \quad (8)$$

where $K(ij)$ is the conditioning set of the node whose conditioned variables are $\{i,j\}$, and the product is taken over all nodes in the vine.

Proof. Re-indexing if necessary, let $\{1,2|3,...,n\}$ denote the constraint of the single node of topmost tree T_{n-1}. Collect all m-descendants of this node containing variable 1. By proposition 2.2; 1 occurs only in the conditioned sets of the m-descendant nodes, and the conditioning set of an m-child is a subset of the conditioning set of its m-parent. By lemma 2.1 variable 1 occurs exactly once with every other variable in the conditioned set of some node. Re-indexing $\{2,..n\}$ if necessary, we may write the constraints of the m-descendants containing 1 of the top node as

$$\{1,2|3,...,n\},\ \{1,3|4,...,n\},\ ...\{1,n-1|n\},\ \{1,n\}.$$

The partial correlations associated with these m-descendant nodes are

$$\rho_{1,2;3,...,n},\ \rho_{1,3;4,...,n},\ ...\rho_{1,n-1;n},\ \rho_{1,n}.$$

and are exactly the terms occurring in (6); hence we may replace the terms in the product on the right hand side of (8) containing these partial correlations by $1 - R^2_{1\{2,...,n\}}$. Note that (6) is invariant under permutation of $\{2,...,n\}$. Remove variable 1 and nodes containing 1; these are just the nodes whose constraints are given above. We obtain the subvine over variables $\{2,...,n\}$. By proposition 2.2; 2 is in the conditioned set of the top node of this subvine. We apply the same argument re-indexing $\{3,...,n\}$ if necessary. With this re-indexing, we may replace the product of terms in (8)

$$(1 - \rho^2_{2,3;4,...,n}),\ (1 - \rho^2_{2,4;5,...,n}),\ ...(1 - \rho^2_{2,n}).$$

by $1 - R^2_{2\{3,...,n\}}$. Proceeding in this way we obtain (7).□

Thus a partial correlation vine represents a factorization of the determinant of the correlation matrix. Following the approach of [21], we would like to change partial correlations to zero while disturbing the empirical partial correlation vine as little as possible. In this case, however, we can first choose the partial correlation vine which best lends itself to this purpose. Moreover, by Theorem 4.1, any such change will be consistent; there is no need for an analogue to the iterative proportional fitting algorithm in the independence graph method. We write

$$-\log(D) = -\sum_{\{i,j\}} \log(1 - \rho^2_{ij;K(ij)}) \quad (9)$$

$$= -\sum_{\{i,j\}} b_{ij;K(ij)} \quad (10)$$

with $\log(1-\rho^2_{ij;K(ij)}) = b_{ij;K(ij)}$. The terms $b_{ij;K(ij)}$ will depend on the regular vine which we choose to represent the second order structure, however the sum of these terms must satisfy (9). We seek a partial correlation vine for which the terms $b_{ij;K(ij)}$ in (9) are "as spread out" as possible. This concept is made precise with the notion of *majorization* [15].

Definition 4.3 *Let x,$y \in \mathbb{R}^n$ be such that $\sum_{i=1}^n x_i = \sum_{i=1}^n y_i$; then x majorizes y if for all $k; k = 1, ..., n$*

$$\sum_{j=1}^{k} x_{(j)} \leq \sum_{j=1}^{k} y_{(j)} \tag{11}$$

where $x_{(j)}$ is the increasing arrangement of the components of x, and similarly for y.

In view of (9) the model inference problem may be cast as the problem of finding a regular vine whose terms $b_{ij;K(ij)}$ are non-dominated in the sense of majorization. In that case, setting those partial correlations equal to zero whose square is smallest will change the determinant as little as possible. Finding non-dominated solutions may be difficult, but a necessary condition for non-dominance can be found by maximizing any Schur convex function.

Definition 4.4 *A function $f : \mathbb{R}^k \to \mathbb{R}$ is Schur convex if $f(x) \geq f(y)$ whenever x majorizes y.*

Schur convex functions have been studied extensively. A sufficient condition for Schur convexity is given by [15].

Proposition 4.1 *If $f : \mathbb{R}^k \to \mathbb{R}$ may be written as $f(x) = \sum f_i(x_i)$ with f_i convex, then f is Schur convex.*

Vine Inference Strategy The following strategy for model inference suggests itself:

1. Choose a Schur convex function $f : \mathbb{R}^{\frac{n(n-1)}{2}} \to \mathbb{R}$;
2. Find a partial correlation vine $\mathcal{V}(n)$ whose vector $b_{ij;K(ij)}$ maximizes f;
3. Set the partial correlations in $\mathcal{V}(n)$ equal to zero for which the terms $b_{ij;K(ij)}$ are smallest;
4. Using the sampling distribution for the determinant, verify that the change in the determinant is not statistically significant.

If the joint distribution is normal, then the sampling distribution of the determinant is given by [6]:

Theorem 4.3 *Let $X = (X_1, ..., X_N)$ be samples of a $n-$dimensional normal vector with sample mean $\overline{X}$, variance V and normalized sample variance*

$$S = N^{-1} \sum_{i=1}^{N} (X_i - \overline{X})(X_i - \overline{X})^T,$$

then

$$\frac{N det(S)}{det(V)} \sim \prod_{i=1}^{N} T_i \tag{12}$$

where $\{T_i\}$ are independent chi-square distributed variables with $N - i + 1$ degrees of freedom.

If the distribution is not joint normal, we could estimate the distribution of the determinant with the bootstrap. Evidently, this strategy depends on the choice of Schur convex function. Searching the set of all partial correlation vines is not easy; at present heuristic constraints are required. We may also consider constrained searches, whereby we require that a given partial correlation be represented in the vine, perhaps with a given value.

Example We consider an example treated in [4]. We have 8 variables corresponding to weather monitoring stations in Europe. The original data has a sample correlation matrix

$$\begin{bmatrix} 1 & .35 & .50 & .49 & .68 & .38 & .50 & .59 \\ & 1 & .79 & .69 & .12 & .64 & .62 & .49 \\ & & 1 & .72 & .18 & .61 & .58 & .43 \\ & & & 1 & .05 & .46 & .47 & .43 \\ & & & & 1 & .33 & .51 & .71 \\ & & & & & 1 & .97 & .77 \\ & & & & & & 1 & .90 \\ & & & & & & & 1 \end{bmatrix}$$

with determinant D satisfying:

$$-log(D) = 11.4406.$$

We wish to 'add independence' so as to obtain a new correlation matrix whose determinant D^* will be larger. Equivalently, $-log(D^*) < -log(D)$. Application of the independence graph method led to setting 17 of the 28 partial correlations in the scaled inverse covariance matrix equal to zero. We find in this case $-log(D^*) = 10.7763$.

To compare this with the vine based approach, we adopt a heuristic search based on maximizing the Schur convex function

$$f(x) = \sum x_i \ln(x_i). \tag{13}$$

The heuristic works as follows.

1. Choose an ordering of the variables.
2. Start with subvine consisting of variables 1 and 2 in the ordering. For $j = 3, \ldots, n$; find the subvine extending the current subvine by adjoining variable $j + 1$, so as to maximize (13). Store the vine obtained for $j = n$.
3. Goto 1.
4. Choose the optimal partial correlation vine maximizing (13) among all those stored.

In general it is not feasible to search all permutations; heuristic search methods or Monte Carlo sampling must be used.

When we set the 17 smallest partial correlations in the optimal vine equal to zero, we find $-log(D^*) = 11.0970$, which is closer to the sample value than that found by the independence graph method. In this sense, the vine method retains the same number of interactions as the independence graph method, while perturbing the sample distribution less.

	2	3	4	5	6	7	8
1	12=0	13;2=0	14;235=0	15;32=0.70	16;235=0	17;235=0	18;2356=0
2		23=0.79	24;53=0	25;3=0	26=0.64	27=0.62	28;6=0
3			34;5=0.72	35=0	36;2=0	37;2=0	38;26=0
4				45=0	46;2351=0	47;2351=0	48;12356=0.49
5					56;23=0	57;23=0.54	58;236=0.78
6						67;12345=0.98	68=0.77
7							78;123456=0.89

Table 1: Partial correlations after setting 17 smallest equal to zero in optimal vine.

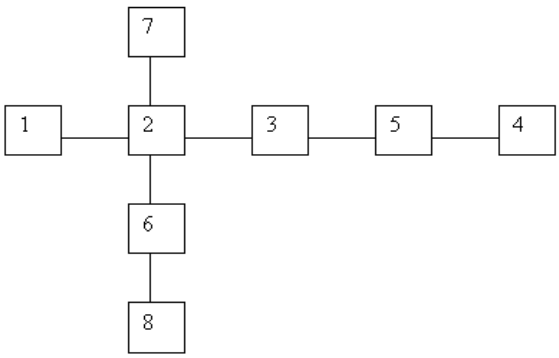

Figure 3: Tree 1 in optimal vine, node 2 has degree 4.

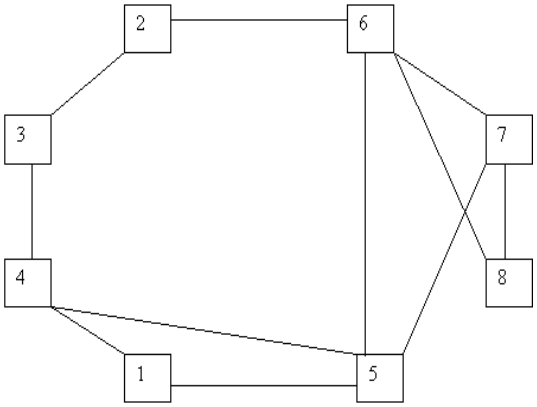

Figure 4: Independence graph from method of [21].

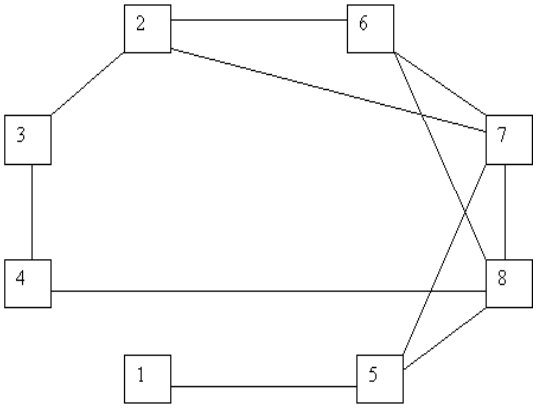

Figure 5: Interaction graph from optimal vine.

Figure 3 shows the first tree in the optimal vine. Note that it is neither a canonical nor D-vine; node 2 has degree 4. Table 1 shows the matrix of partial correlations, in which the 17 partial correlations of smallest magnitude have been set equal to zero. Figure 4 shows the independence

graph obtained by the method of [21]. Figure 5 uses a similar format to show the 11 significant interactions from the vine method. We term this an "interaction graph"; it is not an independence graph in so far as the partial correlations corresponding to edges may not be conditioned on *all* remaining variables.

5 Conclusions

Vines are shown to be a very flexible multivariate dependence model. They enable construction of a joint distribution by specifying the second order structure that has the same number of degrees of freedom as in the correlation matrix, without imposing any algebraic constraints. Moreover, vines can be sampled on the fly with the efficient algorithms presented in this paper. We showed here how to approach the problem of inferring a vine from the data. Vine's inference procedure is in sense similar to the independence graph method. It can perform better then this method but the calculation cost is very extensive. This procedure must still be optimized in terms of applicability in large problems.

References

[1] Dempster A.P. Covariance selection. *Biometrics*, 28:157–175, 1972.

[2] Bedford T.J. and Cooke R.M. Probability density decomposition for conditionally dependent random variables modeled by vines. *Annals of Mathematics and Artificial Integigence*, 32:245–268, 2001.

[3] Bedford T.J. and Cooke R.M. Vines - a new graphical model for dependent random variables. *Ann. of Stat.*, 30(4):1031–1068, 2002.

[4] Callies U., Kurowicka D. and Cooke R.M. Graphical models for the evaluation of multisite temperature forecasts:comparison of vines and independence graphs. *Proc. of ESREL 2003, Safety and Reliability*, 1:363–371, 2003.

[5] Cooke R.M., Waij R. Monte carlo sampling for generalized knowledge dependence with application to human reliability. *Risk Analysis*, 6:335–343, 1986.

[6] Rao C.R. *Linear statistical inference and its applications.* Wiley, New York, 1973.

[7] Lewandowski D. Generalized diagonal band copulae - applications to the vine–copula method, 2004. Presented at the DeMoSTAFI 2004 conference in Quebec, Canada.

[8] Doruet Mari D., Kotz S. *Correlation and Dependence.* Imperial College Press, London, 2001.

[9] Joe H. *Multivariate Models and Dependence Concepts.* Chapman & Hall, London, 1997.

[10] Bojarski J. A new class of band copulas - distributions with uniform marginals. Technical Publication, Institute of Mathematics, Technical University of Zielona Góra, 2001.

[11] Kendall M.G. and Stuart A. *The advenced theory of statistics.* Charles Griffin & Company Limited, London, 1961.

[12] Kiiveri H. and Speed T.P. *Sociological Methodology*, chapter Structural analysis of multivariate data. Jossey Bass, San Francisco, 1984.

[13] Kurowicka D. and Cooke R.M. A parametrization of positive definite matrices in terms of partial correlation vines. *Linear Algebra and its Applications*, 372:225–251, 2003.

[14] Kurowicka D., Cooke R.M. Conditional and partial correlation for graphical uncertainty models. *in Recent Advances in Reldiability Theory, Birkhauser, Boston*, pages 259–276, 2000.

[15] Marshall A.W. and Olkin I. *Inequalities: theory of Majorization and its Applications.* Academic Press, San Diego, 1979.

[16] Meeuwissen A., Bedford T.J. Minimally informative distributions with given rank correlation for use in uncertainty analysis. *J. of Statistical Computation and Simulation*, 57(1-4):143–175, 1997.

[17] Frank M.J. On the simultaneous associativity of f(x,y) and x+y-f(x,y). *Aequationes Math.*, 19:194–226, 1979.

[18] Nelsen R.B. *An Introduction to Copulas.* Springer, New York, 1999.

[19] Cooke R.M. Markov and entropy properties of tree and vines- dependent variables. In *Proceedings of the ASA Section of Bayesian Statistical Science*, 1997.

[20] Ferguson T.S. A class of symmetric bivariate uniform distributions. *Statistical Papers*, 35(1):31–40, 1995.

[21] Whittaker, J. *Graphical Models in applied multivariate statistics.* John Wiley and Sons, Chichester, 1990.

[22] Yule G.U. and Kendall M.G. *An introduction to the theory of statistics.* Charles Griffin & Co. 14th edition, Belmont, California, 1965.