Completion problem with partial correlation vines

D. Kurowicka, R.M. Cooke

March 21, 2005

Abstract

This paper extends the results in [7]. We show that a partial correlation vine represents a factorization of the determinant of the correlation matrix. We show that the graph of an incompletely specified correlation matrix is chordal if and only if it can be represented as an m-saturated incomplete vine, that is, an incomplete vine for which all edges corresponding to membership-descendents (m-descendents for shot) of a specified edge are specified. This enables us to find the set of completions, and also the completion with maximal determinant for matrices corresponding to chordal graphs.
Keywords: correlation, graphs positive definite matrix, matrix completion.
AMS classifications: 60H30, 6209, 62D05, 62H05, 15A48, 15A99

1 Introduction

This paper extends the results in [7]. We show that the graph of an incompletely specified correlation matrix is chordal if and only if it can be represented as an m-saturated incomplete vine, that is, a vine whose all edges corresponding to membership-descendents (m-descendents for short) of a specified edge are specified. We also show that the product of 1 minus the squared partial correlations over all edges of any regular vine equals the determinant of the correlation matrix. Combining with previous results, we have the following picture: a partially specified correlation matrix corresponding to a chordal graph is completable if and only if the partial correlations in its m-saturated vine representation are in the interval $(-1, 1)$, the set of completions are obtained from the set of assignments of values in $(-1, 1)$ to the unspecified edges of the m-saturated vine, and the completion with maximal determinant is obtained by assigning 0 to the unspecified edges.

After briefly reviewing definitions of graphs, regular vines and their properties, we prove that a partial correlation vine represents a factorization of the determinant. We then examine the completion problem for special graphs. Using a result on junction trees, we show that having a chordal graph is equivalent to being representable as an *m-saturated incomplete vine*. This enables us to find the set of completions, and the completion with maximal determinant for matrices corresponding to chordal graphs.

2 Preliminary definitions

2.1 Partial and multiple correlations

A 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 2.1 (Partial correlation)

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

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 [11].

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

Definition 2.2 (Multiple correlation) *The* **multiple correlation** $R_{1\{2,\ldots,n\}}$ *of variables* X_1 *with respect to* $X_2,\ldots,X_n$ *is*

$$1 - R^2_{1\{2,\ldots,n\}} = \frac{D}{C_{11}}$$

where D *is the determinant of the correlation matrix. It is the correlation between* 1 *and the best linear predictor of* X_1 *based on* $X_2,\ldots,X_n$.

In [6] it shown that $R_{1\{2,\ldots,n\}}$ is non negative and satisfies:

$$1 - R^2_{1\{2,\ldots,n\}} \quad = \quad (1-\rho^2_{1n})(1-\rho^2_{1,n-1;n})(1-\rho^2_{1,n-2;n-1,n})\ldots(1-\rho^2_{12;3\ldots n}). \tag{2}$$

It follows from [6], that $R_{1\{2,\ldots,n\}}$ is invariant under permutation of $\{2,\ldots,n\}$ and

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

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

2.2 Graphs

In this subsection we introduce few definitions used in graph theory. Let $G = (N,E)$ be a graph with nodes N and edges E. We say that $G(U)$ is a *subgraph* of a graph G if $G(U)$ is the graph with node set $U \subseteq N$ and with edge set $\{\{uv\} \in E | u,v \in U\}$. We say that $G(U)$ is a subgraph of G *induced* by the vertex set U. U is a *clique* in G when $G(U)$ is a maximal *complete subgraph*, that is a graph in which every pair of vertices is joined. The subset of vertices of G is called complete if it induces a complete subgraph. A *path* of length k between vertices α and β is a sequence $\alpha = \alpha_0,\ldots,\alpha_k = \beta$ of distinct vertices of G such that $(\alpha_{i-1},\alpha_i) \in E$ for all $i = 1,\ldots,k$. A *cycle* of length k is a path of length k in which the end points are identical. A subset $C \subseteq V$ is said to be (α,β)-*separator* if all paths from α to β intersect C. The subset C is said to *separate* A *from* B if it is an (α,β)-separator for every $\alpha \in A$ and $\beta \in B$.

A triple (A,B,C) of disjoint subsets of the vertex set N of an undirected graph G is said to form a *decomposition* of G, or to *decompose* G, if $N = A \cup B \cup C$ and the following two conditions hold:

1. C separates A from B;
2. C is a complete subset of N.

The above definition allows any of sets A, B and C to be empty. If both A and B are non-empty, we say that the decomposition is *proper*.

We say that an undirected graph G is *decomposable* if either

1. it is complete, or;
2. it possesses a proper decomposition (A, B, C) such that both subgraphs $G(A \cup C)$ and $G(B \cup C)$ are decomposable.

The graph G is said to be *chordal* if every cycle of G with length ≥ 4 has a chord, (a *chord* of the cycle C is an edge joining two nonconsecutive nodes of C).

$T = (N, E)$ is a *tree* with nodes N and edges E if E is a subset of unordered pairs of N with no cycle. Hence tree is a special case of chordal graph.

Let $\mathcal{C}$ be a collection of subsets of a finite set N and $\mathcal{T}$ a connected tree with $\mathcal{C}$ at its node set. $\mathcal{T}$ is said to be a *junction tree* if any nonempty intersection $C_1 \cap C_2$ of a pair C_1, C_2 of sets in $\mathcal{C}$ is contained in every node on the unique path in $\mathcal{T}$ between C_1 and C_2. If $\mathcal{C}$ is a family of cliques of a graph G and $\mathcal{T}$ is a junction tree with $\mathcal{C}$ its node set then we say that $\mathcal{T}$ is a *junction tree (of cliques) for the graph* G. The junction tree $\mathcal{T}$ for graph G with k cliques $C_1, C_2, ..., C_k$ has the property that the intersections $S_{ij} = C_i \cap C_j$ between any two neighboring nodes in $\mathcal{T}$ separate graph G [4]. S_{ij} is called *separator* associated with the edge between C_i and C_j.

The following shows a very strong connection between chordality and decomposability of a graph.

Theorem 2.1 *[4] A graph G is decomposable if and only if G is chordal.*

Theorem 2.2 *[4] There exists a junction tree $\mathcal{T}$ of cliques for the graph G if and only if G is chordal.*

3 Vines

Graphical models called *vines* were introduced in [10, 3, 7]. A vine $\mathcal{V}$ on n variables is a nested set of trees $\mathcal{V} = (T_1, ..., T_{n-1})$ 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 3.1 (Regular vine) *$\mathcal{V}$ is a regular vine on n elements if*

1. $\mathcal{V} = (T_1, \ldots, T_{n-1})$,

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.*

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 $j - 1$, $j = 1, ..., n-1$.

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*. More precisely the constraint, the conditioning and the conditioned set of an edge can be defined as follows:

[1]The degree of node is the number of edges attached to it.

Definition 3.2 **1.** *For $e \in E_i, i \leq n-1$*

U_e^* is the **complete union** *of e, that is, the subset of $\{1,\ldots,n\}$ reachable from e by the membership relation.*

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

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

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

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

3. *The* **constraint set** *associated with $e = \{j,k\}$ is*

$$CV_e = \{D_e, C_{e,j}, C_{e,k}\}.$$

Note that for $e \in E_1$, the conditioning set is empty. One can see that the order of an edge is the cardinality of its conditioning set.
For $e \in E_i, i \leq n-1, e = \{j,k\}$ we have $U_e^* = U_j^* \cup U_k^*$.

The following proposition is proved in [3, 7]:

Proposition 3.1 *Let $\mathcal{V} = (T_1, \ldots T_{n-1})$ 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.*

The following definitions provide the vocabulary for studying vines.

Definition 3.3 (m-child; m-descendent) *If node e is an element of node f, we say that e is an* **m-child** *of f; similarly, if e is reachable from f via the membership relation: $e \in e_1 \in \ldots \in f$, we say that e is an* **m-descendent** *of f.*

Lemma 3.1 *For any node K of order $k > 0$ in a regular vine[2], if variable i is a member of the conditioned set of K, then i is a member of the conditioned set of exactly one of the m-children of K, and the conditioning set of an m-child of K is a subset of the conditioning set of K.*

Proof. If the conditioning set of an m-child of K is vacuous, the proposition is trivially true, we therefore assume $k > 1$. Let $K = \{A, B\}$ 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 $k-2$. C_X denote the constraint set of node X, we write $C_K = C_A \cup C_B$. 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 K.□

The edges of a regular vine may be associated with partial correlations, with values chosen arbitrarily in the interval $(-1, 1)$ in the following way:

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

$$\rho_{j,k;D_e}.$$

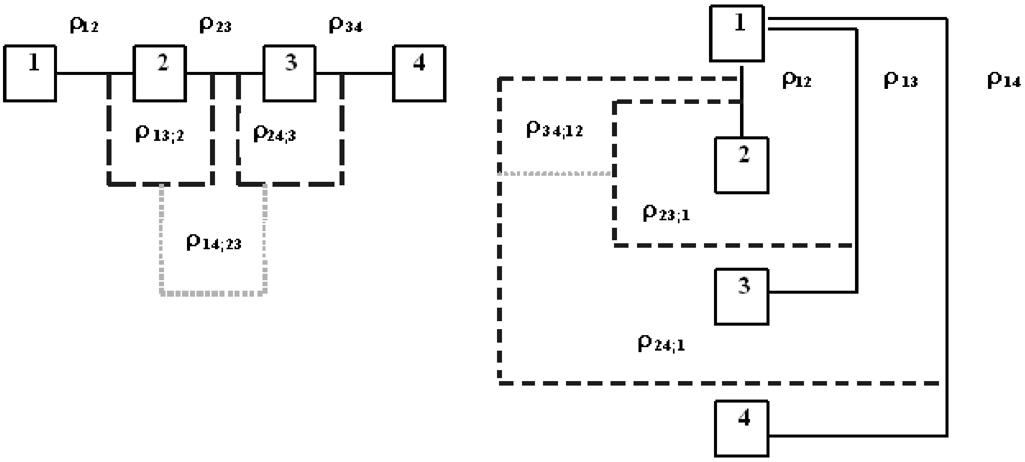

Figure 1: Partial correlations D-vine (left) and canonical vine (right) on 4 variables.

The result is called a *partial correlation vine.*

Theorem 3.1 [3] shows that each such partial correlation vine specification uniquely determines the correlation matrix, and every full rank correlation matrix can be obtained in this way. In other words, a regular vine provides a bijective mapping from $(-1,1)^{\binom{n}{2}}$ into the set of positive definite matrices with 1's on the diagonal.

Theorem 3.1 *For any regular vine on n elements there is a one to one correspondence between the set of $n \times n$ positive definite correlation matrices and the set of partial correlation specifications for the vine.*

All assignments of the numbers between -1 and 1 to the edges of a partial correlation regular vine are consistent, in the sense that there is a joint distribution realizing these partial correlations, and all correlation matrices can be obtained this way.

One verifies that the correlation between ith and jth variables can be computed from the sub-vine generated by the constraint set of the edge whose conditioned set is $\{i, j\}$ using recursive the formulae (1), and the following lemma.

Lemma 3.2 *[7] If $z, x, y \in (-1,1)$, then also $w \in (-1,1)$, where*

$$w = z\sqrt{(1-x^2)(1-y^2)} + xy.$$

A regular vine may thus be seen as a way of picking out partial correlations which uniquely determine the correlation matrix and which are algebraically independent. The partial correlations in a partial correlation vine need not satisfy any algebraic constraint like positive definiteness. The "completion problem" for partial correlation vines is therefore trivial. An incomplete specification of a partial correlation vine may be extended to a complete specification by assigning arbitrary numbers in the $(-1,1)$ interval to the unspecified edges in the vine.

Partial correlation vines have another important property that the product of 1 minus the square partial correlations equals the determinant of the correlation matrix.

[2] Equivalently one can formulate this lemma for edges of $\mathcal{V}$.

Theorem 3.2 *Let D be the determinant of the correlation matrix of variables $1, ..., n$; with $D > 0$. For any partial correlation vine;*

$$D = \prod_{i=1}^{n-1} \prod_{e \in E_i} (1 - \rho^2_{j,k;D_e}) \qquad (4)$$

where $\{j, k\}$ and D_e are conditioned and conditioning sets of e.

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 Lemma 3.1, 1 occurs only in the conditioned sets of the m-descendent nodes, and the conditioning set of an m-child is a subset of the conditioning set of its m-parent. By Proposition 3.1 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 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-descendent nodes are

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

and are exactly the terms occurring in (2). Hence we may replace the terms in the product on the right hand side of (4) containing these partial correlations by $1 - R^2_{1\{2,...,n\}}$. Note that (2) 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 Lemma 3.1, 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 (4)

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

by $1 - R^2_{2\{3,...,n\}}$. Proceeding in this way we obtain (4).□

4 The completion problem

A symmetric real $(n \times n)$ matrix with off-diagonal elements in the interval (-1,1) and with "1"'s on the main diagonal is called a *proto correlation matrix*. Let A be an $n \times n$ partially specified proto correlation matrix such that the unspecified cells are given by index pairs

$$(i_k, j_k), (j_k, i_k), \quad k = 1, ..., K. \qquad (5)$$

We must fill these unspecified elements such that the resulting matrix $B = [b_{ij}]_{i,j=1,...,n}$ is positive definite. Thus we must find a vector $(x_1, ..., x_k)$, such that

$$\begin{aligned} b_{i_k,j_k} &= b_{j_k,i_k} = x_k, \quad k = 1, ..., K, \\ b_{ij} &= a_{ij}, \qquad \text{otherwise} \end{aligned}$$

and B is positive definite.

4.1 Optimization approach

We could approach this problem by trying to find a projection of A on the set of positive definite matrices. However, the constraint of positive definiteness is quite strong. The set of positive definite matrices is not a simple set. Thus, algorithms that search elements of this set are complicated.

The partial correlations specified on a regular vine are algebraically independent and they uniquely determine the correlation matrix. Thus, the partial correlation vine can be seen as an algebraically independent parametrization of the set of correlation matrices.

We thus formulate the completion problem as the following optimization problem. Let A be a partially specified proto-correlation matrix for n variables, let $\mathcal{V}$ be a regular vine on n variables, let x be a vector of partial correlations assigned to the edges of $\mathcal{V}$, and let $B(x)$ be the correlation matrix calculated from $\mathcal{V}$ with x. We then minimize

$$\sum |A_{ij} - B(x)_{ij}|$$

where the sum is over the specified cells of A. If the sum is zero, then A is completable. Notice that the set of vectors x which we must search is simply $(-1,1)^{\binom{n}{2}}$.

4.2 Completion problem for some types of graphs

For some special cases of partially specified matrices the completion problem is quite simple. They are usually discussed from the perspective of the graph corresponding to the matrix.

Definition 4.1 *Let A be an n-by-n partially specified proto correlation matrix. Let $G = (N, E)$ be a graph with nodes $N = \{1, 2, \ldots, n\}$ and edges set E. G is said to be a* **graph for** *A or A is said to* **correspond to** *G (denoted by A_G) if for $i \neq j$, $\{i, j\} \notin E$ implies that a_{ij} and a_{ji} are unspecified.*

Deciding whether a partially specified matrix A can be completed to a positive definite matrix is simple if the graph G for A is a chordal graph. In this case A_G is completable if for every clique K in G the submatrix of A corresponding to elements in K is positive definite [1]. A matrix with this property is called *partial positive definite*. The positive definite completion of A, say B, that has maximal determinant can be obtained. B has a property that unspecified cells of A are equal to zero in the inverse of B [5].

If A corresponds to G which is a cycle then to decide about the completability of A the following condition must be checked [2]. First, for all e in E for which the corresponding element of A is denoted a_e, one must calculate

$$b_e \;=\; \frac{1}{\pi}\arccos(a_e).$$

A is completable if $b = (b_e)_{e \in E}$ satisfy condition

$$\sum_{e \in F} b_e - \sum_{e \in C \setminus F} b_e \leq |F| - 1$$
$$\text{for } C \text{ a cycle in } G, F \subseteq C \text{ with } |F| \text{ odd.}$$

The set of completions for A_G can be found by solving a system of equations [2].

The above conditions separately or coupled are proved to be sufficient also in different cases. For extensive exposition we refer the reader to e.g. [8, 9].

4.3 Completion problem with vines

In analogy to incompletely specified correlation matrices one can define an *incomplete regular vine* and *incomplete partial correlation regular vine* (incomplete vine, and incomplete partial correlation vine for short). An incomplete vine is a regular vine from which some edges have been removed, and an incomplete partial correlation vine is an assignment of partial correlations to the edges of an incomplete vine. An incomplete proto correlation matrix A is analogous to an incomplete partial correlation vine and the graph for A is analogous to the incomplete vine.

The completion problem for special cases of graphs can be solved using incomplete partial correlation vines [7]. We show that chordal graphs correspond to incomplete regular vines with a very simple structure. In this case the set of completions and the completion maximizing the determinant are easily obtained.

Definition 4.2 (m-saturated incomplete vine) *An incomplete vine is* **m-saturated** *if all m-descendants of an edge in its edge set belong to the edge set of the incomplete vine.*

In Figure 2 an m-saturated and non m-saturated D-vines on four variables are shown.

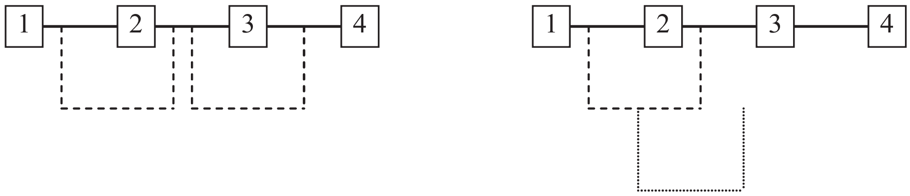

Figure 2: M-saturated (left) and non m-saturated (right) D-vines on 4 variables.

Remark 4.1 *Notice that an m-saturated vine can be easily extended to fully specified vine by adding missing edges in trees of higher order. This extension is not in general unique.*

Lemma 4.1 *Let A be an incomplete proto correlation matrix and let $\mathcal{V}$ be an incomplete vine whose conditioned sets correspond to the specified cells of A. If $\mathcal{V}$ is m-saturated, then A determines a unique partial correlation specification of $\mathcal{V}$.*

Proof. We must show that if ρ_{ij} is specified in A; then the partial correlation $\rho_{ij;D}$ in $\mathcal{V}$ with conditioned set $\{i,j\}$ can be calculated from the specified correlations (we note that these values need not be in $(-1,1)$). The proof is by induction on the cardinality of D. When D has cardinality zero the statement is obvious. The m-saturation property together with regularity means that correlations for the conditioned sets in the subvines over $\{i,D\}$ and $\{j,D\}$ are given. Since $\{i,j\}$ is specified, then the submatrix of A on $\{i,j,D\}$ is fully specified. Hence any partial correlations involving variables in $\{i,j,D\}$ can be calculated. □

The value $\rho_{ij;D}$ calculated in the above proof need *not* be in $(-1,1)$, as indeed A may not be completable. Note also that if the edge $e \in E_{n-1}$ belongs to the edge set of an m-saturated incomplete vine, then there are in fact no missing edges and the vine is complete. An m-saturated incomplete partial correlation vine $\mathcal{V}$ corresponds to a incomplete correlation matrix A if the conditioned sets of $\mathcal{V}$'s edge set correspond to the specified cells of A. In this case A is of a special form. All submatrices corresponding to the constraint sets of edges in $\mathcal{V}$ are fully specified. Hence the constraint sets of the edges in $\mathcal{V}$ of highest order are cliques in the graph G corresponding to A. If an incomplete proto-correlation matrix corresponds to an m-saturated incomplete partial correlation vine, then the latter is not in general unique. It suffices to consider an incomplete correlation matrix in which only cell $\{i,j\}$ is unspecified. Any regular vine whose single edge $e \in E_{n-1}$ has conditioned set $\{i,j\}$ will, upon removing this edge, yield a m-saturated incomplete vine which can be specified by A.

We can define a partial ordering of edges in an m-saturated vine as follows:

$$e_1 \succ e_2 \text{ if } e_2 \text{ is m-descendent of } e_1. \tag{6}$$

For chordal graphs we can construct m-saturated incomplete vines preserving the separation properties of the graph.

Theorem 4.1 *$G = (N, E)$ is chordal if and only if there exists an m-saturated incomplete vine $\mathcal{V}$ on N with conditioned sets E.*

Proof. Let a chordal graph G be given. Since G is chordal then by Theorem 2.2 there exists a junction tree $\mathcal{T}$ with cliques $C_1, ..., C_k$ for G. We construct an m-saturated vine corresponding to $\mathcal{T}$ recursively. One clique in $\mathcal{T}$, say C_1, corresponds to a complete vine on elements of C_1. Let us assume that we have constructed an m-saturated vine for $C_1, ..., C_j$. We show how this m-saturated vine can be extended by adding C_{j+1}, $1 \leq j \leq k-1$. If the intersection of C_{j+1} with $C_1, ..., C_j$ is empty then subvine on elements of C_{j+1} is unconstrained. However, if there exists $1 \leq i \leq j$ such that $S_{i,j+1} = C_i \cap C_{j+1} \neq \emptyset$ then variables in $S_{j+1} = \bigcup_{i=1}^{j} S_{i,j+1}$ have been already placed in the m-saturated vine for $C_1, ..., C_j$. The m-saturated subvine for S_{j+1} is already determined. Since the m-saturated vine can be always extended to complete vine (see remark 4.1) then the subvine for C_{j+1} is constrained to contain the m-saturated subvine for S_{j+1}. For the m-saturated vine the subvines corresponding to cliques $C_1, ..., C_k$ are complete vines and the union of their constraint sets is E.

Conversely, let $\mathcal{V}$ be a m-saturated vine. Let $e_1, e_2, ..., e_k$ be maximum edges of $\mathcal{V}$ with respect to partial ordering defined in (6). Let $C_1, C_2, ..., C_k$ be the constraint sets of $e_1, ..., e_k$, respectively. Let G be a graph corresponding to $\mathcal{V}$. We prove that the graph G is decomposable by induction on the number of maximal edges. If $\mathcal{V}$ is complete (it has one maximum edge) then G is decomposable, otherwise, consider a completion of $\mathcal{V}$ as in remark 4.1. Let the constraint set of the top node e of this completion be $i, j|N \setminus \{i, j\}$. By Lemma 3.1 and Proposition 3.1 i is in the conditioned sets of all m-children of nodes containing i, and never occurs in a conditioning set. It follows that all nodes containing i must be ordered by the m-descendent relation, and this means that the constraint sets of nodes containing i are ordered by the subset relation. Hence only one clique, say C_s can contain i.

Consider the triple $(A, B, C) = (\{i\}, N \setminus C_s, C_s \setminus \{i\})$. Since i belongs only to C_s then the set C separates A from B, moreover since C_s is a clique then C is a complete subset of N. $G(A \cup C)$ is complete hence decomposable and $G(B \cup C)$ is decomposable by the induction step. From Theorem 2.1 G is chordal. □

In [7] it is shown that positive definiteness of a proto correlation matrix can be checked by calculating partial correlations on a regular vine. If all partial correlations on this vine are from the interval (-1,1) then this matrix is positive definite. Hence checking completability of a partially specified proto correlation matrix corresponding to an m-saturated incomplete vine specification $\mathcal{V}$ is quite easy: if all partial correlations obtained via Lemma 4.1 are in $(-1, 1)$, then the matrix is completable. Moreover, the set of completions is easily obtained as the set of assignments of values in $(-1, 1)$ to the unspecified edges of $\mathcal{V}$. It follows from Theorem 3.2 that the completion with maximal determinant is obtained by assigning the value 0 to the unspecified partial correlations. We record this as:

Theorem 4.2 *If A is a partially positive definite correlation matrix then the set of completions of A is the set of correlation matrices computed by assignments of values in $(-1, 1)$ to the unspecified edges to the m-saturated incomplete partial correlation vine corresponding to A. The completion with maximal determinant is the correlation matrix computed by assigning the value 0 to the unspecified edges.*

Example 4.1 *Let us consider the chordal graph G in Figure 3.*

This graph corresponds to the m-saturated vine $m\mathcal{V}$ in Figure 4 (left).

All partial correlations assigned to the edges of $m\mathcal{V}$ can be calculated from specified correlations in the correlation matrix corresponding to G. $m\mathcal{V}$ can be extended to completely specified vine $\mathcal{V}$ in Figure 4 (right). Of course this extension is not unique. Additional edges in $\mathcal{V}$ are presented with dotted lines. From Theorem 3.2 the determinant D of the correlation matrix corresponding to $\mathcal{V}$ can be calculated as

$$D = (1-\rho_{12}^2)(1-\rho_{23}^2)(1-\rho_{24}^2)(1-\rho_{45}^2)(1-\rho_{26}^2)$$

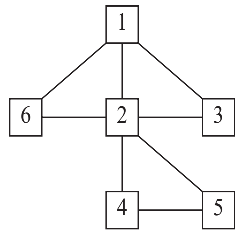

Figure 3: Chordal graph with 6 nodes.

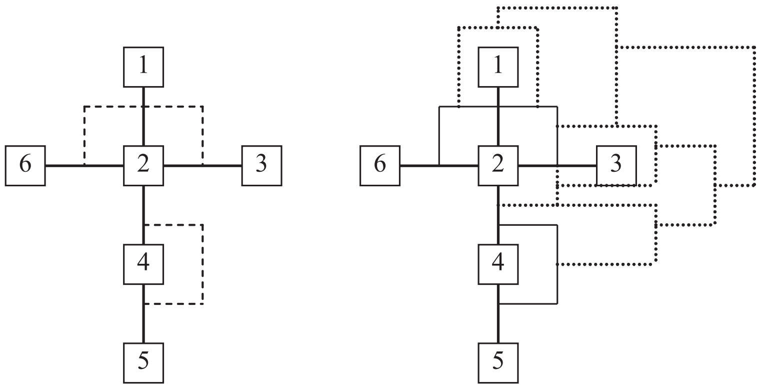

Figure 4: A m-saturated vine (left) and its extension to the complete vine.

$$(1-\rho^2_{13;2})(1-\rho^2_{16;2})(1-\rho^2_{25;4})(1-\rho^2_{34;2})$$
$$(1-\rho^2_{36;12})(1-\rho^2_{14;23})(1-\rho^2_{35;24})$$
$$(1-\rho^2_{46;123})(1-\rho^2_{15;234})(1-\rho^2_{56;1234}).$$

Correlations $\rho_{12}, \rho_{23}, \rho_{24}, \rho_{45}, \rho_{26}, \rho_{13;2}, \rho_{16;2}, \rho_{25;4}$ *are given or can be calculated using recursive formula 1. Other partial correlations can be chosen freely. To obtain the completion with maximum determinant one must take*

$$\rho_{34;2} = \rho_{36;12} = \rho_{14;23} = \rho_{35;24} = \rho_{46;123} = \rho_{15;234} = \rho_{56;1234} = 0.$$

In [7] the partial correlation specification on the canonical vine was used to find the completion of some types of partially specified matrices. In particular the authors showed how to solve the completion problem for A_G where G is a wheel[3].

5 Conclusions

In this paper we have shown how the matrix completion problem can be solved using partial correlation vine. The idea was to simplify the problem by transforming it to the algebraically independent representation of the correlation matrix. Completion problem using the optimization approach becomes then very simple. For special types of matrices, corresponding to chordal graphs, this transformation allows to describe all sets of completions. For other cases however there is no guidance which vine should be used.

[3] A wheel is a graph composed of simple cycle and additional node connected to all nodes in the cycle.

References

[1] Barret W, Johnson C.R and Lindquist M. Determinantal formulae for matrix completions associated with chordal graphs. *Linear Algebra and its Applications*, 121:265–289, 1989.

[2] Barret W, Johnson C.R and Tarazaga P. The real positive definite completion problem for a simple cycle. *Linear Algebra and its Applications*, 192:3–31, 1993.

[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] Cowell R.G., Dawid A.P, Lauritzen S.L. and Spiegelhalter D.J. *Probabilistic Networks and Expert Systems.* Statistics for Engineering and Information Sciences. Springer- Verlag, New York, 1999.

[5] Grone R., Johnson C.R, Sà E.M and Wolkowicz H. Positive definite completions og partial hermitian matrices. *Linear Algebra and its Applications*, 58:109–124, 1984.

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

[7] 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.

[8] Laurent M. Polinomial instances of the positive semidefinite and euclidean distance matrix completion problems. SIAM Journal on Matrix Analisis and its Applications, 1993.

[9] Laurent M. The real positive semidefinite completion problem for series-parallel graphs. *Linear Algebra and its Applications*, 253:347–366, 1997.

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

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