A Constraint View of IBD Graphs

Rina Dechter, Dan Geiger and Elizabeth Thompson

Donald Bren School of Information and Computer Science
University of California, Irvine, CA 92697

1 Introduction

The report provides a constraint processing view of Identity by decent (IBD) graphs and reformulate some aspects of the work in [4] from a constraint perspective.

A Bayesian network model of a linkage task can be decomposed into a mixed network that has a constraint network portion and a Bayesian network portion [3]. It includes selector variables (or inheritance variables) that determine the flow of genes from parents to children along the chromosome which are described by probabilistic dependencies. We also have the prior probabilities over the founders. These two sets of probabilistic functions can be regarded as a Bayesian subnetwork. The rest of the dependencies in the model are deterministic and can be viewed as constraints. In a multimarker model, whenever the selector variables are conditioned on, we have a collection of independent small mixed networks, one for each locus. The constraint portion of each locus-based mixed network, can be processed by a constraint propagation algorithms. In particular, it can be processed by arc and path-consistency [1]. We recently observed that if we apply path-consistency *symbolically*, (i.e., when the values of the variables (the alleles) are propagated symbolically), then we get a tighter constraint network restricted to the founder variables which is equivalent (identical) to the *identity by decent (IBD)* graph [4].

This equivalent constraint subnetwork together with the probabilistic subnetwork is an equivalent model at each locus, conditioned on the selector values. These networks can be far smaller than the original ones , and when provided with the actual evidence (the alleles assigned to the typed individuals), they can often (always?) be solved efficiently. In particular, enumerating the set of all consistent alleles associated with the founder variables can be done in output linear time. Clearly

if the number of consistent founder assignments for a locus-based IBD constraint network is small, computing the probability of evidence conditioned on the selectors s and over all markers can be accomplished more effectively using the tighter mixed IBD networks than when processing the original one.

The main virtue of the IBD graph seems to be that it changes only locally from one locus to the next, and only for selectors that represent recombinations [4]. In other words, the IBD constraint network along the chromosome will mimic recombination and will be more a function of the total number of recombinations rather than the number of markers. In this report we propose to use the mixed network view of the IBD graph to facilitate constraint-based techniques to advance ideas and goals in linkage analysis and haplotype computations.

Section 2 provides general background on on mixed probabilistic and deterministic networks and on its use for formulating the linkage analysis task. Section 3 provides the formulation of IBD graphs.

2 Background: the Mixed network of Linkage analysis

2.1 Definitions

DEFINITION **1 (constraint network)** *A constraint network $\mathcal{CN}$ is a triple $\mathcal{CN} = \langle X, D, C \rangle$, where $X = \{X_1, ..., X_n\}$ is a set of variables associated with a set of discrete-valued domains $D = \{D_1, ..., D_n\}$ and a set of constraints $C = \{C_1, ..., C_r\}$. Each constraint C_i is a pair $\langle S_i, R_i \rangle$ where R_i is a relation $R_i \subseteq D_{S_i}$ defined on a subset of variables $S_i \subseteq X$ and D_{S_i} is the Cartesian product of the domains of variables S_i. The relation R_i denotes all tuples of D_{S_i} allowed by the constraint. The projection operator π creates a new relation, $\pi_{S_j}(R_i) = \{x \mid x \in D_{S_j}$ and $\exists y, y \in D_{S_i \setminus S_j}$ and $x \cup y \in R_i\}$, where $S_j \subseteq S_i$. Constraints can be combined with the join operator $\bowtie$, resulting in a new relation, $R_i \bowtie R_j = \{x \mid x \in D_{S_i \cup S_j}$ and $\pi_{S_i}(x) \in R_i$ and $\pi_{S_j}(x) \in R_j\}$.*

DEFINITION **2 (constraint satisfaction problem, satisfiability)** *The constraint satisfaction problem (CSP) defined over a constraint network $\mathcal{C} = \langle \mathcal{X}, \mathcal{D}, \mathcal{C} \rangle$, is the task of finding a solution, that is, an assignment of values to all the variables $x = (x_1, ..., x_n), x_i \in D_i$, such that $\forall C_i \in C, \pi_{S_i}(x) \in R_i$. The set of all solutions of the constraint network $\mathcal{C}$ is $sol(\mathcal{C}) = \bowtie \mathcal{R}_j$. When the variables are propositional, having values "0" and "1" and the constraints are boolean clauses we have the special case of a cnf formula and the satisfiability task.*

Graphical models can accommodate both probabilistic and deterministic information. Probabilistic information typically associates a strictly positive number with an assignment of variables, quantifying our expectation that the assignment may be realized. The deterministic information has a different semantics, annotating assignments with binary values, either *valid* or *invalid*. The mixed network allows probabilistic information expressed as a belief network and a set of constraints to co-exist side by side and interact by giving them a coherent umbrella meaning.

DEFINITION **3 (mixed networks)** *Given a belief network $\mathcal{B} = \langle \mathbf{X}, \mathbf{D}, \mathbf{G}, \mathbf{P} \rangle$ that expresses the joint probability $P_{\mathcal{B}}$ and given a constraint network $\mathcal{R} = \langle \mathbf{X}, \mathbf{D}, \mathbf{C} \rangle$ that expresses a set of solutions $\rho(\mathcal{R})$ (or simply ρ), a mixed network based on $\mathcal{B}$ and $\mathcal{R}$ denoted $\mathcal{M}_{(\mathcal{B},\mathcal{R})} = \langle \mathbf{X}, \mathbf{D}, \mathbf{G}, \mathbf{P}, \mathbf{C} \rangle$ is created from the respective components of the constraint network and the belief network as follows. The variables $\mathbf{X}$ and their domains are shared, (we could allow non-common variables and take the union), and the relationships include the CPTs in $\mathbf{P}$ and the constraints in $\mathbf{C}$. The mixed network expresses the conditional probability $P_{\mathcal{M}}(\mathbf{X})$:*

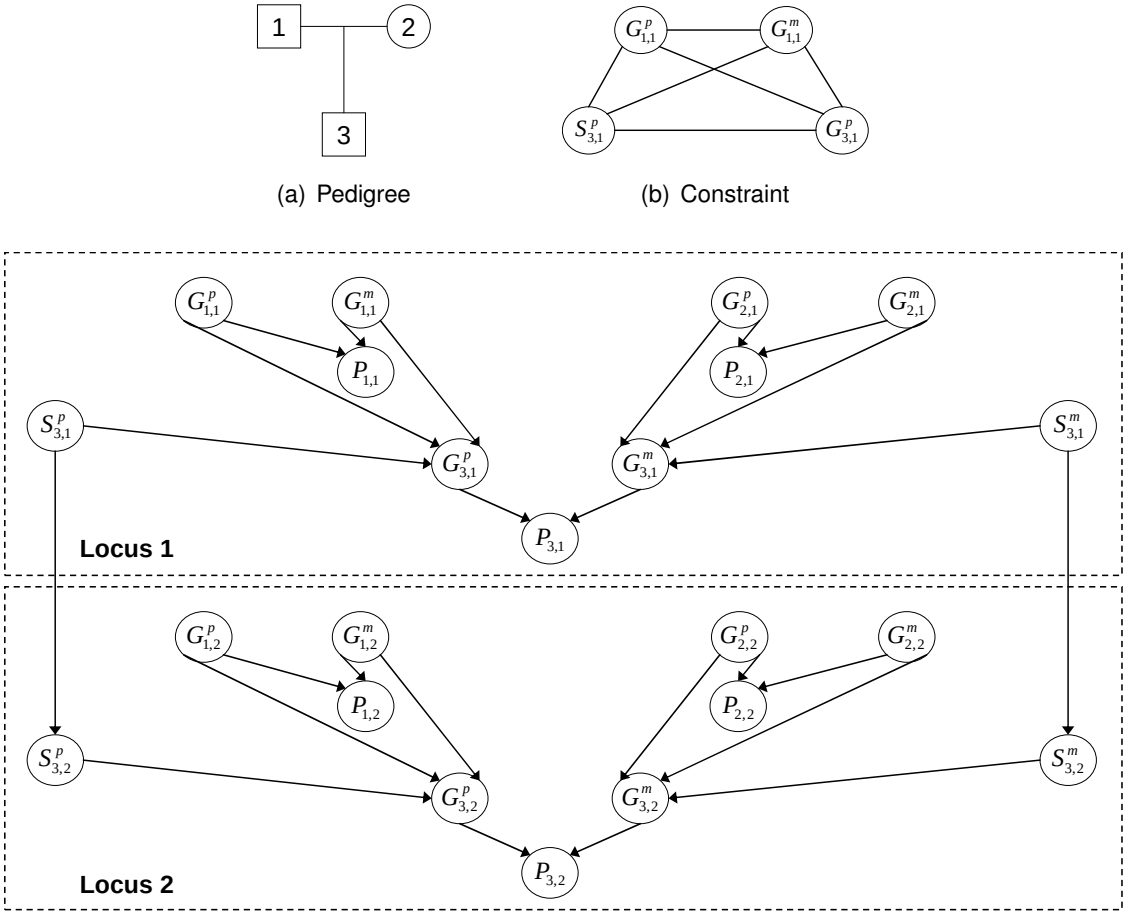
$$P_{\mathcal{M}}(\bar{x}) = \begin{cases} P_{\mathcal{B}}(\bar{x} \mid \bar{x} \in \rho), & if \ \ \bar{x} \in \rho \\ 0, & otherwise. \end{cases}$$

Clearly, $P_{\mathcal{B}}(\bar{x} \mid \bar{x} \in \rho) = \frac{P_{\mathcal{B}}(\bar{x})}{P_{\mathcal{B}}(\bar{x} \in \rho)}$. By definition, $P_{\mathcal{M}}(\bar{x}) = \prod_{i=1}^{n} P(x_i \mid \bar{x}_{pa_i})$ when $\bar{x} \in \rho$, and $P_{\mathcal{M}}(\bar{x}) = 0$ when $\bar{x} \notin \rho$. When clarity is not compromised, we will abbreviate $\langle \mathbf{X}, \mathbf{D}, \mathbf{G}, \mathbf{P}, \mathbf{C} \rangle$ by $\langle \mathbf{X}, \mathbf{D}, \mathbf{P}, \mathbf{C} \rangle$ or $\langle \mathbf{X}, \mathbf{P}, \mathbf{C} \rangle$.

Queries over Mixed Networks: Belief updating, MPE and MAP queries can be extended to mixed networks straight-forwardly. They are well defined relative to the mixed probability distribution $P_{\mathcal{M}}$. Since $P_{\mathcal{M}}$ is not well defined for inconsistent constraint networks, we always assume that the constraint network portion is consistent, namely it expresses a non-empty set of solutions. An additional relevant query over a mixed network is to find the probability of a consistent tuple relative to $\mathcal{B}$, namely determining $P_{\mathcal{B}}(\bar{x} \in \rho(\mathcal{R}))$. It is called *CNF or Constraint Probability Evaluation (CPE)*. Note that the notion of evidence is a special type of constraint. For linkage analysis the primary query of interest is to compute the probability of evidence.

2.2 Modeling linkage analysis by mixed network

We describe next the problem of genetic linkage analysis [**?**], which is usually formulated as a belief network, but can be represented as a mixed network to leverage the deterministic information abundantly present.



(a) Pedigree (b) Constraint

(c) Mixed network

Figure 1: Genetic linkage analysis

Genetic linkage analysis is a statistical method for mapping genes onto a chromosome, and determining the distance between them. This is very useful in practice for identifying disease genes. Without going into the biology details, we briefly describe how this problem can be modeled as a reasoning task in a mixed network.

Figure 1(a) shows the simplest pedigree, with two parents (denoted by 1 and 2) and an offspring (denoted by 3). Square nodes indicate males and circles indicate females. Figure 1(c) shows the usual belief network that models this small pedigree for two particular loci (locations on the chromosome). There are three types of variables, as follows. The G variables are the genotypes (the values are the specific alleles, namely the forms in which the gene may occur on the specific locus), the P variables are the phenotypes (the observable characteristics). Typically these are

evidence variables, and for the purpose of the graphical model they take as value the specific unordered pair of alleles measured for the individual. The S variables are selectors (taking values 0 or 1). The upper script p stands for paternal, and the m for maternal. The first subscript number indicates the individual (the number from the pedigree in 1(a)), and the second subscript number indicates the locus. The interactions between all these variables are indicated by the arcs in Figure 1(c).

Due to the genetic inheritance laws, many of these relationships are actually deterministic. For example, the value of a selector variable determines the genotype variable. Formally, if a is the father and b is the mother of x, then:

$$G^p_{x,j} = \begin{cases} G^p_{a,j}, & if \;\; S^p_{x,j} = 0 \\ G^m_{a,j}, & if \;\; S^p_{x,j} = 1 \end{cases}$$

and,

$$G^m_{x,j} = \begin{cases} G^p_{b,j}, & if \;\; S^m_{x,j} = 0 \\ G^m_{b,j}, & if \;\; S^m_{x,j} = 1 \end{cases}$$

The CPTs defined above are in fact deterministic, and can be captured by a constraint, whose constraint graph is depicted graphically in Figure 1(b). The only real probabilistic information is defined by the CPTs between selector variables and the prior probabilities of the founders, namely the individuals having no parents in the pedigree. Figure 2 provides the mixed network formulation of a founder variable (top of Figure), on the bottom left we have the Bayes subnetwork that consists of three independent variables and on the right there is a constraint subnetwork. Figure 3 describes the 3 member family formulation as a mixed network.

Genetic linkage analysis is an example of a belief network that contains many deterministic or functional relations that can be exploited as constraints. The typical reasoning task is equivalent to computing the probability of the evidence, or to maximum probable explanation.

3 A constraint network view of IBD graphs

To describe a constraint formulation of IBD graphs I will use the description of ibd graphs in section 2.3 of [4]). Paper [4] provides a description of earlier work (by Sobel and Lang (1996) and by Kruglyak et. al. (1996)) of what is called *distinct-genome-lable* (DGL) graph that allow the computation of $P(Y_{\bullet,j}|S_{\bullet,j})$. Given an assignment to the inheritance variables $S_{\bullet,j}$ in a particular marker locus j, and a set of distinct labels for each of the two genomes of each founder, the DGL that can

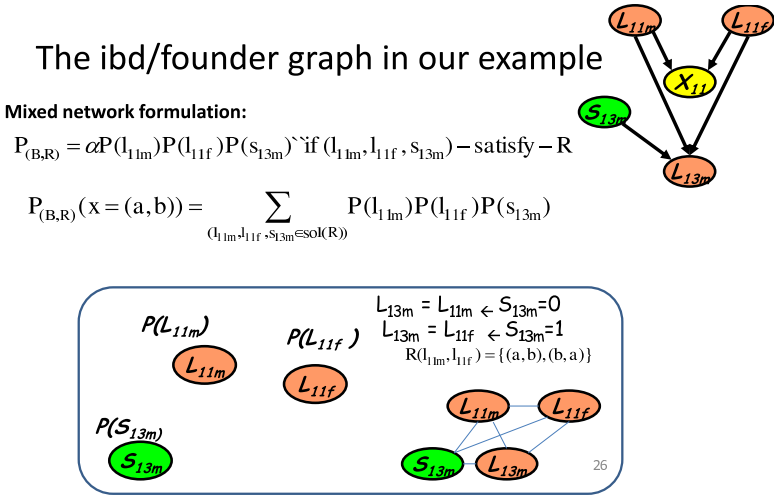


Figure 2: A non-founder mixed network

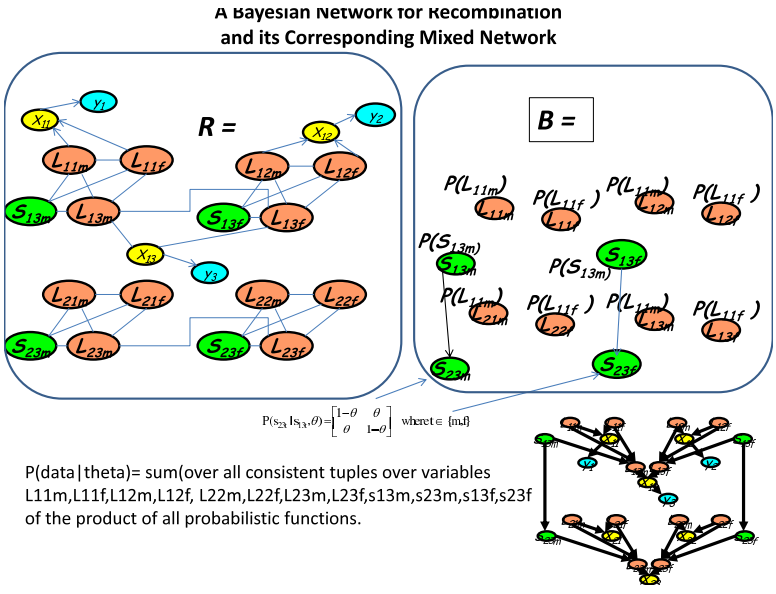


Figure 3: A mixed network for recombination

be assigned to each individual in the pedigree are known and unique. Therefore when we are given a set of observed non-founders and an assignment to the selectors we can deduce the founders that contributed the genomes to each observed type and therefore can infer the founders actual (DGL's) labels.

DEFINITION **4 (ibd graphs [4])** *Given a Bayesian network $\mathcal{B}$, modeling a linkage instance and given an assignment to the inheritance variables $S_{\bullet,j}$ in a particular marker locus j and given a set of observed types, the ibd graph of $\mathcal{B}$ is defined by a set of nodes which correspond to the founders' maternal and paternal genoms. Two nodes having opposing gender are connected if there exists an observed type whose labels are ibd with the maternal and paternal labels of its neighbor founder nodes according to the model $\mathcal{B}$. The arc will be labeled by all the observed types that satisfy this condition. The graph captures the ibd that is implied by the selectors and the observed non-founders in the given network $\mathcal{B}$.*

Formally: Given $\mathcal{B}$, and its implied mixed network defined over a set of founders variables $\mathcal{F} = \{F_1^m, F_1^p, ...F_i^m, F_i^p ....F_k^m, F_k^p\}$, a set of individual non-founders $I = \{I_1, ...I_n\}$ and a subset of non-founders that are typed $O = \{O_1^m, O_1^p, ..., O_r^m, O_r^p\}$, the IBD graph *$\mathcal{G}_{j,s}$ for locus j and selector $S_{\bullet,j} = s_{\bullet,j}$ is defined by $\mathcal{G} = (\mathcal{F}, \mathcal{E}, O)$ where $\mathcal{F}$ is a subset of $F_\bullet^m, F_\bullet^p$, $E \subseteq F_\bullet^m \times F_\bullet^p$ and O is the set of labels. We say that an IBD graph is consistent with a model $\mathcal{B}$ iff for any two nodes that are connected and labeled by $O_t \in O$, the labels observed in O_t can be provably inherited from the neighboring founders labels (given the selectors' assignments), according to model $\mathcal{B}$.*

Consider for example Figure 4. In this graph nodes 2 and 9 are connected and labeled by A. This means that the two labels observed at A are inherited from the maternal locus of one founder (labeled 9) and the paternal label of the founder denoted by 2. The example does not indicate what are the selector values but we can figure it out from the description. In principle we can have 2N nodes if there are N founders. However only a subset of those founder variables that are relevant to the observed non-founders in the pedigree, are included in the ibd graph. We next define the ibd-graph with a constraint network that represents the same information.

DEFINITION **5 (ibd constraint network)** *Given a Bayesian network model $\mathcal{B}$ having a set of founders $\mathcal{F} = \{F_1^m, F_1^p, ...F_i^m, F_i^p ....F_k^m, F_k^p\}$ a set of individual non-founders $I = \{I_1, ...I_n\}$ and a subset of non-founders that are typed $O = \{O_1^m, O_1^p, ..., O_r^m, O_r^p\}$, and given a consistent IBD graph $\mathcal{G}$, $\mathcal{G} = (\mathcal{F}, E, O)$, the* IBD constraint network *of the IBD $\mathcal{G}$, denoted $CONS(\mathcal{G}) = (\mathcal{X}, \mathcal{D}, \mathcal{C})$ has a set of variables $X = \mathcal{F} \cup O$ (namely, $X = \{F_1^m, F_1^p, ...F_k^m, F_k^p, O_1^m, O_1^p, ..., O_r^m, O_r^p\}$). The domains of all non-observed variables (founders and non-founders) are all the possible alleles at locus j. Each arc in the IBD graph implies a set of constraints on the ibd-constraint as follows: If arc $(F_i^m, F_k^p) \in E$*

and is labeled O_l, then there is a constraint over $F_i^m, F_k^p, O_l^m, O_l^p$ that forces that the maternal and paternal labels in O_l are identical to one of the values of F_i^m and F_k^p. Using Boolean constraints of disjunction and inequality, these constraints can be expressed as:

$$F_i^m = O_l^m \vee O_l^p, \;\; F_k^p = O_l^m \vee O_l^p, \;\; F_i^m \neq F_k^p \tag{1}$$

The actual alleles for the typed individuals are modeled separately by na *evidence domain constraints*. By definition, consistent founder assignments of the IBD graph correspond to solutions of the IBD-constraint network in conjunction with its evidence domain constraints.

Example 1 *Consider the example in Figure 5 of [4]. In defining next the IBD constraint network of the ibd graph in Figure 4 we will stay with variable names of founders being numbers and variable names of observed types being alphabetical. The domains constraints associated with evidence will be be lower alphabetical. The variables of the constraint network are*

$$X = \{1,2,3,4,5,6...18,A,B,C,D,E,F,G,H,J,K,L,V,U,W\}$$

The domains of the numerical variables are the possible alleles at that locus. The constraints associated with each arc in the ibd graph are:

$$2 = A^m \vee A^p, \;\; 9 = A^m \vee A^p, \;\; 2 \neq 9 \tag{2}$$

$$2 = B^m \vee B^p, \;\; 13 = B^m \vee B^p, \;\; 2 \neq 13, \;\; 2 = J^m \vee J^p, \;\; 13 = J^m \vee J^p, \tag{3}$$

$$2 = G^m \vee G^p, \;\; 4 = G^m \vee G^p, \;\; 2 \neq 4 \tag{4}$$

$$13 = D^m \vee D^p, \;\; 4 = D^m \vee D^p, \;\; 13 \neq 4 \tag{5}$$

$$13 = E^m \vee E^p, \;\; 6 = E^m \vee E^p, \;\; 13 \neq 6 \tag{6}$$

$$6 = C^m \vee C^p \tag{7}$$

$$6 = H^m \vee H^p, \;\; 15 = H^m \vee H^p, \;\; 6 \neq 15 \tag{8}$$

$$15 = L^m \vee L^p, \;\; 17 = L^m \vee L^p, \;\; 15 \neq 17 \tag{9}$$

$$15 = F^m \vee F^p, \;\; 4 = F^m \vee F^p, \;\; 15 \neq 4 \tag{10}$$

In this example we assume the following observations. Non-founders A, B, J are all observed to have a_1a_4*, type G has* a_1a_6*, type D has* a_4a_6*, type E has* a_4a_2*, C has* a_2a_2*, F has* a_3a_6*, H has* a_2a_3*, and L has* a_1a_3*. This will be modeled as evidence constraints in the evidence constraint network in the form of unary constraints that restricts the variable domains. The evidence constraints are: (*$D_{A\bullet}$ *stands for the pair of constraints:* D_{A^m} *and* D_{A^p} *)*

$D_{A\bullet} = D_{B\bullet} = D_{J\bullet} = \{a_1, a_4\}$,
$D_{G\bullet} = \{a_1, a_6\}$
$D_{D\bullet} = \{a_4, a_6\}$,
$D_{E\bullet} = \{a_4, a_2\}$
$D_{C\bullet} = \{a_2, a_2\}$
$D_{F\bullet} = \{a_3, a_6\}$
$D_{H\bullet} = \{a_2, a_3\}$
$D_{L\bullet} = \{a_1, a_3\}$

For example $D_{D\bullet} = \{a_4, a_6\}$ *stands for the constraints:* $D^m = a_4 \vee a_6$, $D^p = a_4 \vee a_6$ *and* $D^m \neq D^p$*. (or* $D^m = a_4 \rightarrow D^p = a_6$ *and* $D^p = a_4 \rightarrow D^m = a_6$ *)*

In this example there is a single solution which can be obtained by applying arc-consistency once the IBD constraint subproblems is assigned the actual values (the specific alleles observed for the typed individuals.) From this information we can infer that the label of 2 is a_1*; labels 9, and 13 are* a_4*; 4 is* a_6*; 6 is* a_2*; 15 is* a_3*, and 17 is* a_1*. The probability of this set of labels is* $q_1^2 q_2 q_3 q_4^2 q_6$*.*

3.1 Deriving the IBD constraint networks from the input mixed network

We next show that the locus-based IBD constraint subnetwork can be inferred through path and arc-consistency followed by a removal of irrelevant variables.

A mixed network of a linkage analysis task (decomposed from its Bayes network as described earlier) yields a collection of locus-based mixed networks, one per locus, defined by all its locus variables. This locus-based networks (which ignores the transition dependencies between the selectors of successive loci, expresses a probability distribution at each locus. By definition, the probability of

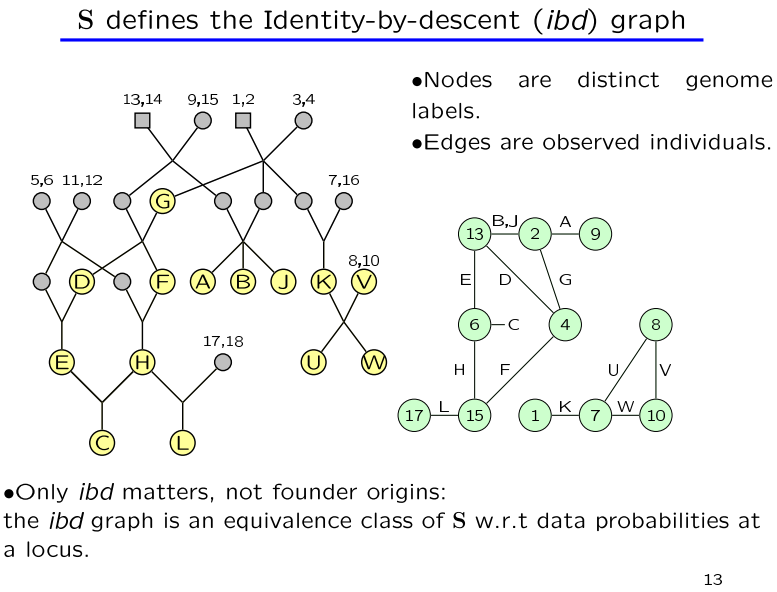


Figure 4: An example of pedigree with its associated IBD graph

any consistent tuple, conditioned on the selectors, is proportional over the $\mathcal{F}$ and O variables to the product of the marginal probabilities in the Bayesian network since the Bayesian network portion is a set of independent variables. In our example instance there is only one tuple of founder variable which is consistent with the observations namely: $(2 = a_1; 9 = a_4, 13 = a_4; 4 = a_6; 6 = a_2; 15 = a_3, 17 = a_1)$, its probability is indeed $q_1^2 q_2 q_3 q_4^2 q_6$.

The constraint subnetwork within any locus mixed network, conditioned on its selectors can be processed by arc and path-consistency in a symbolic manner without using the specific evidence yielding an equivalent constraint network. When its variables are restricted to the founder variables only it becomes far smaller. We will show that this path-consistent network restricted to the relevant founder variables is identical to the IBD constraint network defined earlier. In other words, the IBD constraint networks can be obtained by applying path-consistency over the original set of constraints. For a definition of the application of path-consistency and arc-consistency see [1].

DEFINITION **6 (Path-consistency)** *Given a network of constraints $\mathcal{R} = (X, D, C)$ and given a set of evidence nodes $Y = y$ for $Y \subseteq X$, we define $R' = PC(R, y)$ as the network obtained by applying path-consistency and arc-consistency to $\mathcal{R} \wedge Y = y$. The network R' is defined on the same set of variables as R. The restriction of a network $\mathcal{R}$ to a subset of variables Z, is denoted $\mathcal{R}_Z$.*

Proposition 1 *Given a mixed network model $\mathcal{M}_{j,s} = (\mathcal{B}_{j,s}, \mathcal{R}_{j,s})$ at a locus j,and an assignment to the selector variables $S = s$ of a linkage problem. Let $G_{j,s}$ denotes its IBD graph conditioned*

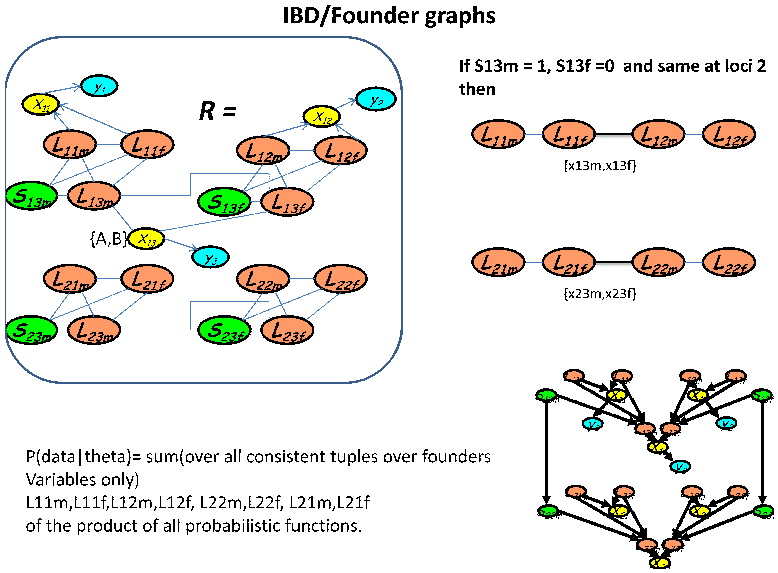


Figure 5: A mixed network represented by the IBD graph

on $S = s$, then the IBD constraint network $CONS(G_{j,s})$, is identical to the path and arc-consistent network derived from $\mathcal{M}_{j,s}$. Namely $CONS(G_{j,s}) = R'_F$ where $R' = PC(\mathcal{R}_{j,s})$.

Proof I think it is correct but need to prove.

Hypothesis:

1. The IBD constraint graphs are always tractable and yield all solutions in output polynomial time.
2. Does applying path and arc-consistency on the IBD-restricted constraint network yield the minimal domains and constraints.

Figure 5 shows the original constraint network at a locus (on the left) and the derived IBD constraint network (on the right).

Computing the probability of evidence. Paper [4] demonstrates how to compute the probability of trait data Z. The paper notes the difference when computing the probability of marker data (Y) and the probability of computing trait data Z. The trait data, given assignments to the inheritance variables, depends only on the marker data to the left and right and on the trait model (phenotype given genotype). Likewise computing the probability of the trait given the selectors can be accomplished over the derived mixed networks of the trait locus (including the trait model itself) which consists of the derived ibd constraint network and the Bayesian networks (that has a collection of disconnected probabilistic tables).

Rephrasing within the mixed network formulation, the probability of the evidence conditioned

on the selectors at locus j is the product of probabilities over all assignments of founder variables that are consistent with the evidence in the corresponding IBD graph. If the number of consistent founder assignments is small, computing the probability of evidence conditioned on the selectors and computing the probability of evidence over all markers can be accomplished along the mixed networks more effectively. However if we have many markers, as is the case in snps data, or if we have complex diseases computation may still be difficult.

The main virtue of the IBD graph seems to be that it changes only locally from one locus to the next, and only for selectors that represent recombinations. In other words, the IBD constraint problem along the chromosome will mimic recombination and will be more a function of the total number of recombinations rather than the number of markers.

3.2 Moving from one locus to the next capturing recombination: sporadic thoughts

Assume now that in addition to the selector variables we add persistence variables $Q_{\bullet,j}$ where $Q_{\bullet,j} = 0$ if there is no change moving from $S_{\bullet,j}$ to $S_{\bullet,(j+1)}$. In other words $Q_{\bullet,j} = |S_{\bullet,j} - S_{\bullet,j+1}|$. The number of changes due to recombination moving from one locus to the next is the number of 1's in this Q delta function. We can now add another auxiliary variable that counts the number of recombinations moving from one locus to the next called $M_{\bullet,j} = \sum_i Q_{\bullet,i,j}$. We can identify markers of interest as those where some recombination could occur, namely those for whom M is greater than 0.

The number of different IBD graphs can be significantly smaller than the number of different selector combinations along the chromosome. We can also assume that since snps are so close, there could be only a single recombination for a single individual moving in from one snp to the next and that for some small interval no recombination occurs. This seems to be what is argued in [4] in section 2.4. This is also consistent with assumptions made in Geiger et. al.'s work on handling SNP data http://cbl-hap.cs.technion.ac.il/superlink-snp/. Section 2.4 also demonstrate how the IBD graph can change along the chromosome due to recombination.

Compiling IBD constraint graphs. The collection of IBD constraint graphs along the chromosome can be compiled and allows a far more manageable computation. One option we may consider is to use AND/OR multi-valued decision diagrams [2]. Another option is proposed and carried out in a recent paper [**?**].

Extension of the mixed network view to inference of identity by descent on two chromosome of the same individual as discussed in Section 3 of [4] could also illuminate the computational aspect

and can bring in additional constraint processing and general graphical models ideas. In particular, modeling LD along the chromosome can be captured through additional (HMM like) transition probabilities between founder variables. Finally extension analyzing chromosome of population, where each individual is captured by its IBD graph along the chromosome are likely to yield far more informative answers. The mixed probabilistic and constraint-based view of this data can bring to bear both advanced computation developed in the constraint community and graphical model communities.

References

[1] R. Dechter. *Constraint Processing*. Morgan Kaufmann Publishers, 2003.

[2] R. Mateescu and R. Dechter. Compiling constraint networks into and/or multi-valued decision diagrams. In *Constraint Porgramming (CP2006)*, 2006.

[3] Robert Mateescu and Rina Dechter. Mixed deterministic and probabilistic networks. *Ann. Math. Artif. Intell.*, 54(1-3):3–51, 2008.

[4] E. A. Thompson. Analysis of data on related individuals through inference of identity by decent. In *Technical report 539, Dept of Statistics, University of Washington*, August, 2008.