Toward Normative Expert Systems: Part II Probability-Based Representations for Efficient Knowledge Acquisition and Inference

David E. Heckerman
Departments of Computer Science and Pathology
University of Southern California
HMR 204, 2025 Zonal Ave
Los Angeles, CA 90033
<dheck@sumex-aim.stanford.edu>

Bharat N. Nathwani
Department of Pathology
University of Southern California
HMR 204, 2025 Zonal Ave
Los Angeles, CA 90033

To appear in *Methods of Information in Medicine,* 1992

Abstract

We address practical issues concerning the construction and use of decision-theoretic or *normative expert systems* for diagnosis. In particular, we examine Pathfinder, a normative expert system that assists surgical pathologists with the diagnosis of lymph-node diseases, and discuss the representation of dependencies among pieces of evidence within this system. We describe the belief network, a graphical representation of probabilistic dependencies. We see how Pathfinder uses a belief network to construct differential diagnoses efficiently, even when there are dependencies among pieces of evidence. In addition, we introduce an extension of the belief-network representation called a *similarity network*, a tool for constructing large and complex belief networks. The representation allows a user to construct independent belief networks for subsets of a given domain. A valid belief network for the entire domain can then be constructed from the individual belief networks. We also introduce the partition, a graphical representation that facilitates the assessment of probabilities associated with a belief network. Finally, we show that the similarity-network and partition representations made practical the construction of Pathfinder.

Keywords: Expert systems, diagnosis, probability theory, decision theory, artificial intelligence, belief networks, similarity networks, partitions, pathology

1 Introduction

Decision-theoretic or *normative expert systems* have not become commonplace because they have been difficult to build and use. In this article, however, we introduce several representations that facilitate the construction and use of normative expert systems for diagnosis. These representations are based on the *belief network* [1, 2], a graphical representation of uncertain knowledge.

We describe extensions to the belief-network representation in the context of Pathfinder, a normative expert system that assists surgical pathologists with the diagnosis of lymph-node diseases [3, 4, 5]. This medical application is an excellent testbed in which to investigate practical issues concerning normative expert systems. The domain is large: More than 60 diseases can invade the lymph node (25 benign diseases, 9 Hodgkin's lymphomas, 18 non-Hodgkin's lymphomas, and 10 metastatic diseases). In addition, there are approximately 130 morphologic, clinical, laboratory, immunologic, and molecular-biologic features that are relevant to the diagnosis of lymph-node diseases. For a detailed description of Pathfinder, see the companion of this article.

Most normative expert systems constructed in the 1960s and 1970s made the inaccurate assumptions that (1) diseases are mutually exclusive and exhaustive, and (2) all features are conditionally independent, given each disease. For the diagnosis of lymph-node diseases, the assumption that diseases are mutually exclusive is appropriate, because co-occurring diseases almost always appear in different lymph nodes or in different regions of the same lymph node. Also, the large scope of Pathfinder makes reasonable the assumption that the set of diseases is exhaustive. The assumption of conditional independence, however, is inaccurate. For example, given certain diseases, finding that mummy cells are present increases greatly the chances that classic Sternberg-Reed cells will be present. Thus, in building Pathfinder, we concentrated on the problem of representing and reasoning with conditionally dependent features. Similarly, in this article, we concentrate on this issue. Nonetheless, the representations described here also facilitate the construction of expert systems for diagnostic problems where multiple diseases may coexist.

2 Belief Networks

The *belief network* is a graphical knowledge representation that rigorously describes any probabilistic-inference problem, yet has a human-oriented qualitative structure that facilitates communication between the expert and the probabilistic model.[1] Several researchers have developed and studied belief networks, although they have used various names for this representation such as *causal nets* [6, 7], *probabilistic cause–effect models* [8], *Bayesian belief networks* and *causal networks* [2, 9, 10, 11, 12], *probabilistic causal networks* [13], and *knowledge maps* [14].

A belief network is a directed acyclic graph that contains nodes that represent uncertain variables, and arcs that represent dependencies among those variables. Figure 1 shows a belief network for the problem of distinguishing ordinary nodular sclerosing Hodgkin's

[1] In this article, we address only the representation of probabilistic-inference problems with belief networks. An extension of belief networks called *influence diagrams* [1] can represent any decision problem.

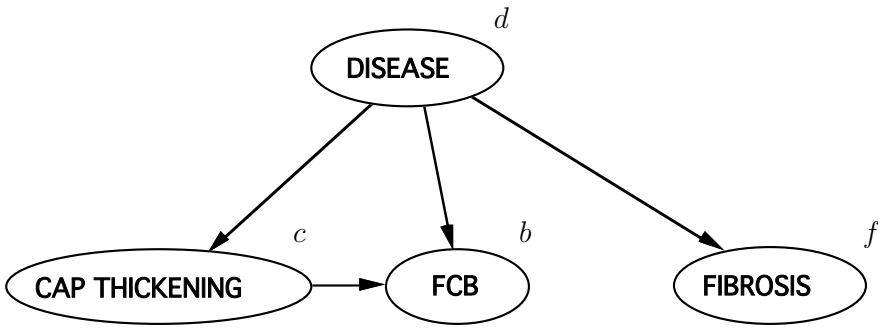
disease from the cellular phase of nodular sclerosing Hodgkin's disease. The node DISEASE represents the two possible diseases, and the nodes CAP THICKENING (capsule thickening), FCB (prominent fibrocollagenous bands), and FIBROSIS (prominent fibrosis) represent the features that are relevant to the discrimination of these two diseases. For presentation purposes, we sometimes shall use the lower-case letters d, c, b, and f to represent the variables DISEASE, CAP THICKENING, FCB, and FIBROSIS, respectively.

Each node in the belief network is associated with a set of mutually exclusive and exhaustive *instances.* We shall denote an instance of a variable by subscripting that variable. For the belief network of Figure 1, the node DISEASE has instances d_{NS} and d_{CP} (ordinary nodular sclerosing Hodgkin's disease and cellular phase of nodular sclerosing Hodgkin's disease); the node CAP THICKENING has instances $c_{<5}$, c_{5-10}, c_{11-20}, and $c_{>20}$ (<5, $5-10$, $11-20$, and >20 small-lymphocyte diameters); the node FCB has instances b_- and b_+ (absent and present); and the node FIBROSIS has instances f_- and f_+ (absent and present).

In a belief network, an arc from node x to node y reflects an assertion by the builder of that network that the probability distribution for y may depend on the the instances of x. We say that the x *conditions* y. For example, in Figure 1, the arcs from the disease node to the feature nodes reflect the expert's belief that the probability of observing a particular instance for each feature may depend on the disease that is present. In addition, the arc from CAP THICKENING to FCB reflects the expert's assertion that the probability distribution for FCB may depend on whether or not there is capsule thickening, even when the identity of the disease is known. Conversely, the lack of arcs in a belief network reflect assertions of conditional independence. In Figure 1, there is no arc between CAP THICKENING and FIBROSIS nor is there an arc between FCB and FIBROSIS. The lack of these arcs encode the expert's assertion that FIBROSIS is conditionally independent of CAP THICKENING and FCB, given the identity of the patient's disease. Later in this section, we examine in more detail the assertions of conditional independence represented by a belief network.

Each node in a belief network is associated with a set of probability distributions. In particular, a node has a probability distribution for every instance of its conditioning nodes. For example, in Figure 1, FIBROSIS is conditioned by DISEASE. Thus, FIBROSIS has two probability distributions (shown below the belief network in Figure 1): $p(f|d_{\text{NS}}, \xi)$, the probability distribution for observing fibrosis given that a patient has ordinary nodular sclerosing Hodgkin's disease, and $p(f|d_{\text{CP}}, \xi)$, the distribution for observing fibrosis given that a patient has the cellular phase of nodular sclerosing Hodgkin's disease. The symbol ξ denotes the background knowledge of the expert who provides the probabilities. Similarly, CAP THICKENING has two probability distributions. In contrast, FCB is conditioned by both DISEASE and CAP THICKENING. Consequently, this node has eight distributions corresponding to the instances where DISEASE is d_{NS} or d_{CP}, and where CAP THICKENING is $c_{<5}$, c_{5-10}, c_{11-20}, or $c_{>20}$. Finally, DISEASE has only one distribution—the *prior probability distribution* of disease—because it is not conditioned by any nodes.

In general, the construction of a belief network is straightforward. First, the builder of the network orders the variables. Second, the *joint probability distribution* over the set of variables is expanded using this ordering and the product rule of probability theory. The joint probability distribution over a set of variables is the collection of probabilities for each instance of that set. For example, given the ordering over n variables $x^1, \ldots, x^n$, we expand



$p(\, d_{\text{NS}} \mid \xi\,) = 0.95$
$p(\, d_{\text{CP}} \mid \xi\,) = 0.05$

$p(\, c_{<5} \mid d_{\text{NS}}, \xi\,) = 0.01$
$p(\, c_{5\text{-}10} \mid d_{\text{NS}}, \xi\,) = 0.1$
$p(\, c_{11\text{-}20} \mid d_{\text{NS}}, \xi\,) = 0.4$
$p(\, c_{>20} \mid d_{\text{NS}}, \xi\,) = 0.49$

$p(\, c_{<5} \mid d_{\text{CP}}, \xi\,) = 0.01$
$p(\, c_{5\text{-}10} \mid d_{\text{CP}}, \xi\,) = 0.3$
$p(\, c_{11\text{-}20} \mid d_{\text{CP}}, \xi\,) = 0.4$
$p(\, c_{>20} \mid d_{\text{CP}}, \xi\,) = 0.29$

$p(\, b_{+} \mid c_{<5}, d_{\text{NS}}, \xi\,) = 0.05$
$p(\, b_{+} \mid c_{5\text{-}10}, d_{\text{NS}}, \xi\,) = 0.30$
$p(\, b_{+} \mid c_{11\text{-}20}, d_{\text{NS}}, \xi\,) = 0.40$
$p(\, b_{+} \mid c_{>20}, d_{\text{NS}}, \xi\,) = 0.55$
$p(\, b_{+} \mid c_{<5}, d_{\text{CP}}, \xi\,) = 0.0$
$p(\, b_{+} \mid c_{5\text{-}10}, d_{\text{CP}}, \xi\,) = 0.0$
$p(\, b_{+} \mid c_{11\text{-}20}, d_{\text{CP}}, \xi\,) = 0.0$
$p(\, b_{+} \mid c_{>20}, d_{\text{CP}}, \xi\,) = 0.0$

$p(\, f_{+} \mid d_{\text{NS}}, \xi\,) = 0.0$
$p(\, f_{+} \mid d_{\text{CP}}, \xi\,) = 0.05$

Figure 1: A belief network for the discrimination of ordinary nodular sclerosing Hodgkin's disease from cellular phase nodular sclerosing Hodgkin's disease. The features relevant to this diagnostic problem are CAP THICKENING (capsule thickening), FCB (prominent fibrocollagenous bands), and FIBROSIS (prominent fibrosis). The arcs from the disease node to the feature nodes reflect the expert's belief that the likelihood of observing each feature may depend on the disease that has manifested in the lymph node. The arc from CAP THICKENING to FCB represents the expert's assertion that the probability of FCB may depend on whether or not there is capsule thickening, given disease. Conversely, the lack of arcs from CAP THICKENING and FCB to FIBROSIS represent the expert's belief that FIBROSIS is conditionally independent of the other two features, given disease. The probability distributions associated with each node are shown below the belief network (see the manuscript for a description of the notation). The probabilities for nodes b and f that are not shown can be computed from the sum rule of probability. For example, $p(b_{-}|c_{<5}, d_{\text{NS}}, \xi) = 1 - p(b_{+}|c_{<5}, d_{\text{NS}}, \xi) = 0.95$.

the joint probability distribution over these variables as follows:

$$p(x^1, \ldots, x^n|\xi) = p(x^1|\xi)\ p(x^2|x^1, \xi)\ \ldots\ p(x^n|x^1, \ldots, x^{n-1}, \xi) \tag{1}$$

Note that Equation 1 is a set of equations: one equation for each instance of the variables. Third, the builder of the network makes assertions of conditional independence that simplify the terms in the expansion. Finally, the expert draws the belief network, given the simplified expansion of the joint probability distribution. In particular, for every variable x, the expert draws an arc to x from each node that conditions x in the simplified expansion.

Let us use this construction method to build the belief network in Figure 1. First, we list the variables in the order d, c, b, and f. Second, we expand the joint probability distribution over these ordered variables to obtain

$$p(d, c, b, f|\xi) = p(d|\xi)\ p(c|d, \xi)\ p(b|c, d, \xi)\ p(f|d, c, b, \xi) \tag{2}$$

Third, we make assertions of conditional independence that simplify the terms on the right-hand side of Equation 2. In this case, we assert

$$p(f|d, c, b, \xi) = p(f|d, \xi) \tag{3}$$

Combining Equations 2 and 3, we obtain the simplified expansion of the joint probability distribution

$$p(d, c, b, f|\xi) = p(d|\xi)\ p(c|d, \xi)\ p(b|c, d, \xi)\ p(f|d, \xi) \tag{4}$$

Finally, we draw the belief network in Figure 1 by examining each term in the expansion of Equation 4. Because d conditions the distribution for c, we draw an arc from d to c; because c and d condition the distribution for b, we draw arcs from c and d to b; and because d conditions the distribution for f, we draw an arc from d to f.

The probability distributions in the expansion of Equation 4 are exactly the distributions associated with the nodes in the belief network of Figure 1. Therefore, this belief network and these probability distributions determine a unique joint probability distribution over the variables d, c, b, and f. In general, a belief network and the probability distributions associated with the nodes in the network determine a unique joint probability distribution over the variables in that network.

Probabilistic inference is the computation—via the rules of probability—of one set of probabilities from another set. Given a joint probability distribution over a set of variables, we can compute any conditional probability that involves those variables. For example, suppose we want to compute the probability that a patient has nodular sclerosing Hodgkin's disease, given that fibrocollagenous bands are present. Applying the rules of probability, we have

$$\begin{aligned} p(d_{\mathrm{NS}}|f_+, \xi) &= \frac{p(d_{\mathrm{NS}}, f_+|\xi)}{p(f_+|\xi)} \\ &= \frac{\sum_{c_j, b_k}\ p(d_{\mathrm{NS}}, c_j, b_k, f_+|\xi)}{\sum_{d_i, c_j, b_k}\ p(d_i, c_j, b_k, f_+|\xi)} \end{aligned}$$

where d_i, c_j, and b_k denote arbitrary instances of the variables d, c, and b, respectively.

Thus, given a belief network for some domain, we can perform any probabilistic inference in that domain by constructing the joint probability distribution from the belief network, and by applying the rules of probability directly to this joint probability distribution. Such computations, however, are often intractable. Fortunately, researchers have developed algorithms for probabilistic inference that exploit the assertions of conditional independence encoded in a belief network [15, 16, 17, 12, 18]. For each of these algorithms, computation time and memory requirements decrease as the number of conditional-independence assertions increases.

A belief network may represent assertions of conditional independence that are not explicitly made by the expert when he constructs the network. Such assertions follow logically from the assertions made by the expert. By identifying additional assertions of conditional independence, we can check the assertions made by the expert, and simplify probabilistic inference. To identify such assertions, we need the following definitions. The *underlying graph* of a belief network is an undirected graph obtained from the belief network by replacing every arc with an undirected edge. A *trail* in a belief network is a sequence of arcs that form a cycle-free path in the underlying graph. A node is a *head-to-head* node along a trail, if there are two consecutive arcs along the trail that both point to the node. For example, b is a head-to-head node along a trail from c to d in Figure 1. A trail is *activated* by a set of nodes Z, if (1) every head-to-head node along the trail either is in Z or has a descendant in Z, and (2) every other node along the trail is outside Z. The sets of variables X and Y are *d-separated* by Z, if no trail from X to Y is activated by Z. For example, both trails from c to f are activated by the empty set. Consequently, c and f are not d-separated by the empty set. In addition, neither trail from c to f is activated by d. Consequently, c and f are d-separated by d.

Pearl states without proof that, if Z d-separates X from Y, then X and Y are conditionally independent, given Z [16]. Verma and Pearl prove this result [10]. In addition, Geiger and Pearl prove that the assertions of conditional independence determined by this d-separation criterion are the only assertions that follow logically from those assertions of conditional independence made explicitly by the network builder. That is, they prove that any other valid assertions of conditional independence are a consequence of the particular probabilities assigned to the network; such assertions are not a consequence of the network structure. For example, because c and f are d-separated by d in the belief network of Figure 1, we know that c and f are conditionally independent, given d. In contrast, because c and f are not d-separated by the empty set, we cannot conclude from this network that c and f are independent.

The belief network for Pathfinder is shown in Figure 2. We describe how this network was created in the next section. To perform inference in this network, we use a special case of the algorithm described by Lauritzen and Spiegelhalter [12]. The special-case algorithm is extremely efficient, because it takes advantage of the fact that many of the arcs in the network emanate from the disease node. On MS-DOS hardware with a 25 megahertz 486 processor and math coprocessor, Pathfinder can construct or update any differential diagnosis in less than 1 second. For a detailed description of this algorithm, see Heckerman [5] or Suermondt et al. [20].

Figure 2: The complete belief network for Pathfinder. The node DISEASE contains more than 60 lymph-node diseases. The appendix contains a key to the feature and disease abbreviations. (Taken with permission from D. Heckerman, *Probabilistic Similarity Networks,* MIT Press, Cambridge, MA, 1991.)

3 Similarity Networks: The Construction of Belief Networks

A belief network simplifies knowledge acquisition by exploiting a fundamental observation about the ability of people to assess probabilities. Namely, a belief network takes advantage of the fact that people can make assertions of conditional independence much more easily than they can assess numerical probabilities [1, 16]. In using a belief network, a person first builds the graph that reflects his assertions of conditional independence, and only then does he assess the probabilities underlying the graph. Thus, a belief network helps a person to decompose the construction of a joint probability distribution into the construction of a set of smaller probability distributions. This decomposition does not sacrifice a precise probabilistic representation nor the need to make erroneous assumptions of conditional independence. If an expert believes that—for example—CAP THICKENING and FCB are conditionally dependent, he can represent this dependency explicitly. On the other hand, if he believes that the features are conditionally independent, he can represent this assertion. In either case, a joint probability distribution over the variables in the domain can be constructed.

Unfortunately, this decomposition does not make practical the construction of the joint probability distribution for extremely large domains. In fact, we were unable to construct directly the belief network for Pathfinder, shown in Figure 2; we were overwhelmed by the number of conditional-independence assertions that we had to consider. Fortunately, however, we developed a representation, called a *similarity network*, that allowed us to decompose the construction of this belief network into a set of tasks of manageable size [21, 22, 5].

A similarity network consists of a similarity graph and a collection of local belief networks. A *similarity graph* is an undirected graph whose vertices (nodes) represent the mutually exclusive diseases, and whose edges connect diseases that an expert considers to be similar or difficult to discriminate in practice. Figure 3 shows the similarity graph for Pathfinder. The edge between INTERFOLLICULAR HD (interfollicular Hodgkin's disease) and MIXED CELLULARITY HD (mixed-cellularity Hodgkin's disease), for example, reflects the expert's opinion that these two diseases are often mistaken for each other in practice.

Associated with each edge in a similarity graph is a *local belief network*. The local belief network for an edge is a belief network that contains only those features that are relevant to the discrimination of the two diseases that are connected by that edge. The local belief networks are typically small, because the disease pairs for which they are constructed are similar. For example, the belief network in Figure 1 is the local belief network for the edge between CELLULAR PHASE NSHD (cellular phase nodular sclerosing Hodgkin's disease) and NODULAR SCLEROSIS HD (ordinary nodular sclerosing Hodgkin's disease) in the similarity graph. The local belief network contains only the features CAP THICKENING, FCB, and FIBROSIS. Thus, the expert believes that only these features are relevant to the discrimination of these two types of nodular sclerosing Hodgkin's disease.[2]

Given a similarity graph and all its associated local belief networks for a given domain, we can construct a single belief network for the entire domain—called the *global belief network*—

[2]The actual local belief network contains additional features. This example was simplified for illustrative purposes.

Figure 3: The similarity graph for Pathfinder. The nodes in the graph represent the mutually exclusive diseases that can manifest in a lymph node. Edges connect diseases that the expert considers to be similar. (Taken with permission from D. Heckerman, *Probabilistic Similarity Networks,* MIT Press, Cambridge, MA, 1991.)

with a simple procedure. In particular, we construct the *graph union* of all the local belief networks. The operation of graph union is straightforward. The nodes in the graph union of a set of graphs is the simple union of the nodes in the individual graphs. Similarly, the arcs in the graph union of a set of graphs is the simple union of the arcs in the individual graphs. That is, a node (or arc) appears in the graph union, if and only if there is such a node (or arc) in at least one of the individual graphs. We constructed the Pathfinder belief network in Figure 2 with this procedure.

Under relatively weak conditions, this construction of the global belief network is *sound* [5]. That is, any joint probability distribution that satisfies the assertions of conditional independence implied by the local belief networks also satisfies the assertions of conditional independence implied by the global belief network. Thus, the similarity-network representation greatly facilitates the construction of large belief networks. In particular, the representation allows an expert to decompose the task of building a large belief network into modular and relatively small subtasks.

Several important features of the similarity-network representation are discussed elsewhere [22, 5]. For example, similarity networks can be extended to include local belief networks for sets of diseases that contain two or more elements. Essentially, we need only to replace the similarity graph with a similarity hypergraph.[3] The representation also can be used in situations where diseases are not mutually exclusive.

A similarity network derives its power from its ability to represent assertions of conditional independence that are not conveniently represented in an ordinary belief network. To illustrate such an assertion, let variable d represent the mutually exclusive and exhaustive diseases $d_1, d_2, \ldots, d_n$. Further, let $d_\subseteq$ denote a subset of these diseases. If d and feature f are independent, given that any one of the elements of $d_\subseteq$ is present, we say that f is *independent of the subset* $d_\subseteq$. Formally, we have

$$p(d_j|f_i, d_\subseteq, \xi) = p(d_j|d_\subseteq, \xi) \tag{5}$$

for all instances f_i of variable f, and for all diseases d_i in $d_\subseteq$. In Equation 5, the set $d_\subseteq$, which conditions both probabilities, denotes the disjunction of its elements. Using Bayes' theorem, we can show that a feature f is independent of the subset $d_\subseteq$, if and only if

$$p(f_i|d_j, \xi) = p(f_i|d_k, \xi) \tag{6}$$

for all pairs $d_j, d_k \in d_\subseteq$, and for all instances f_i of feature f. We call the form of conditional independence represented by Equations 5 and 6 *subset independence*.

Although we cannot easily encode assertions of subset independence in a belief network, we can naturally represent such assertions in a similarity network. In particular, if we omit the feature f from the local belief network for the diseases d_j and d_k, then we are asserting that f is independent of the subset $\{d_j, d_k\}$. In the next section, we examine how to exploit subset independence for probability assessment.

[3] A hypergraph consists of nodes and edges among node sets of arbitrary size.

4 Similarity Networks and Partitions: The Assessment of Probabilities in a Belief Network

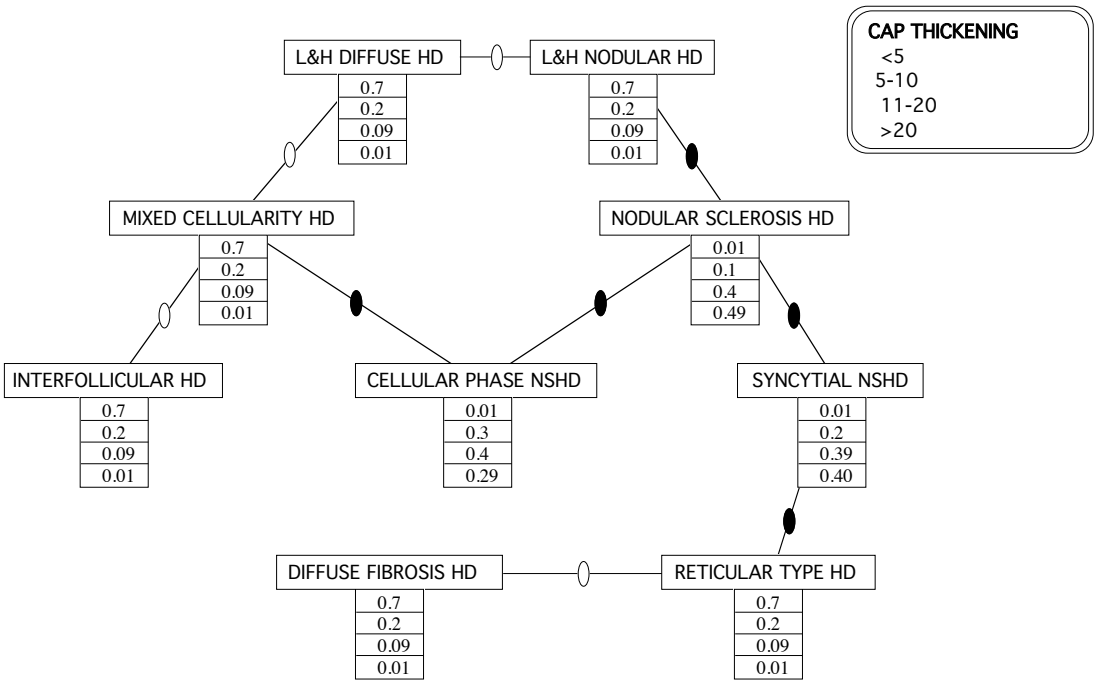
In Section 2, we saw that each node in a belief network is associated with a set of probability distributions. In Figure 1 we represented these distributions simply as a table of numbers. We can, however, represent such distributions in a similarity network. For example, consider the feature CAP THICKENING (thickening of lymph-node capsule). In the global belief network, Figure 2, this feature is conditioned only by DISEASE. Thus, we need to assess the probability distribution for CAP THICKENING, given each disease. Figure 4(a) shows how we can represent these assessments using the Pathfinder similarity graph. In the figure, only the portion of the similarity graph for Hodgkin's diseases is shown. To simplify the presentation, we shall restrict our attention to these diseases in the remainder of this discussion. The rounded rectangle labeled with the feature name contains the mutually exclusive and exhaustive instances for the feature: less than 5, 5–10, 11–20, and greater than 20 small-lymphocyte diameters. The four numbers under each disease are the probability distribution for the feature given that disease. For example, the probability that CAP THICKENING is 11–20 small-lymphocyte diameters, given NODULAR SCLEROSIS HD, is 0.4.

A black oval on an edge in the similarity graph indicates that the feature CAP THICKENING is present in the local belief network corresponding to that edge. Conversely, a white oval on an edge indicates that this feature is absent from that local belief network. As shown in the figure, when a feature is omitted from a local belief network, the conditional probability distributions on either side of an edge are equal. This observation follows from Equation 6 and from the fact that any feature omitted from a local belief network must be independent of the subset consisting of the two diseases associated with that local belief network. Consequently, for the feature CAP THICKENING, we need to assess probability distributions given only L&H NODULAR HD, NODULAR SCLEROSIS HD, CELLULAR PHASE NSHD, SYNCYTIAL NSHD, and RETICULAR TYPE HD.

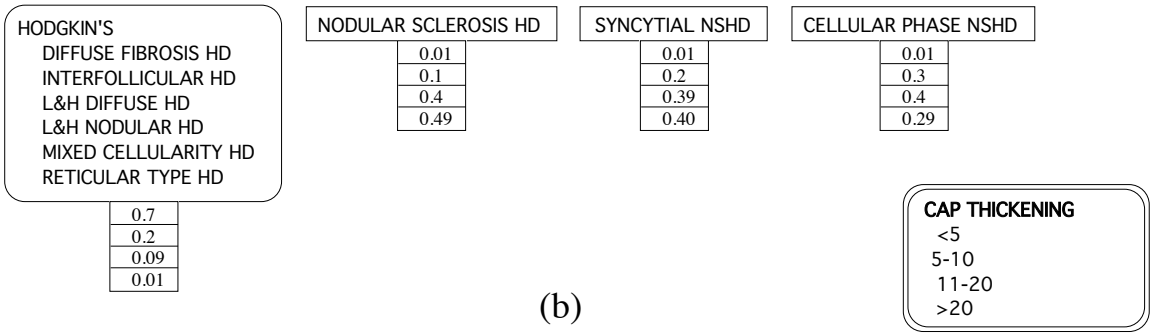
A problem with this approach to probability assessment is illustrated in Figure 4(a). Specifically, the probability distributions for the feature CAP THICKENING given INTERFOLLICULAR HD and DIFFUSE FIBROSIS HD are equal. Because we did not connect these diseases in the similarity graph, however, the equality of these distributions remained hidden until the expert assessed the actual probabilities. We can remedy this difficulty by composing a local belief network for every pair of diseases. For domains such as Pathfinder's that contain many diseases, however, this alternative is impractical.

Alternatively, we can construct a *partition* of the diseases for each feature to be assessed. In composing a partition for a given feature, we place each disease into one and only one set; we place two or more diseases in the same set, thereby forming subset $d_{\subseteq}$, only if the feature is independent of subset $d_{\subseteq}$. After composing the partition for a given feature, we assesses probability distributions for the feature, given each disease. Given Equation 6, however, we need to assess only one probability distribution for each set in the partition.

A partition for the feature CAP THICKENING is shown in Figure 4(b). In this partition, the diseases are divided into four sets: the singleton sets containing NODULAR SCLEROSIS HD, SYNCYTIAL NSHD, and CELLULAR PHASE NSHD, and the set labeled HODGKIN'S that contains the remaining diseases. The partition reflects the assertion that the feature CAP



(a)



(b)

Figure 4: (a) Probability assessment using a similarity network. The probability distributions for the feature CAP THICKENING, given the various types of Hodgkin's disease, are shown. The rounded rectangle labeled with the feature name contains the mutually exclusive and exhaustive instances for the feature: < 5, $5 - 10$, $11 - 20$, and > 20 small-lymphocyte diameters. The numbers below each disease node are the probability distribution for the feature given that disease. A white oval on an edge indicates that the feature is absent in the corresponding local belief network. Conversely, a black oval indicates that the feature is present in the local belief network. Distributions bordering an edge with a white oval must be equal. (b) Probability assessments for the same feature using a partition. In this representation, an expert needs to assess only one probability distribution for each set of diseases.

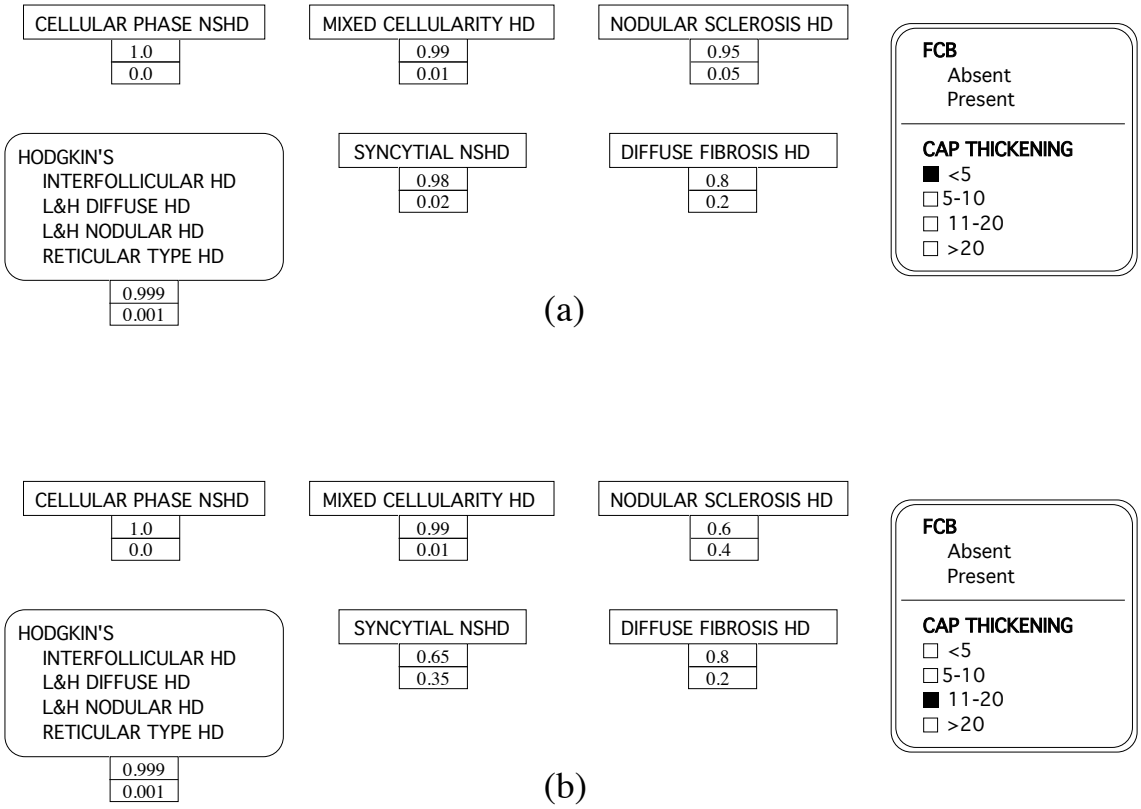


Figure 5: (a) The partition for for FCB (prominent fibrocollagenous bands) given that CAP THICKENING is less than five small-lymphocyte diameters. (b) The partition for for FCB given that CAP THICKENING is between 11 and 20 small-lymphocyte diameters. Note that, in both of these partitions, the subset of diseases named HODGKIN'S is different from the subset of diseases with the same name in Figure 4(b).

THICKENING is independent of the subset HODGKIN'S. That is, if the expert knew that the true disease was in the set HODGKIN'S, then his observation of the status of the lymph-node capsule would not change his relative probabilities of the diseases in that set. Consequently, we need to assess only four probability distributions. These distributions, shown below the sets in Figure 4(b), are the same as those shown in Figure 4(a). By using this partition, we uncover an additional equality among the distributions for CAP THICKENING before we assess probabilities; we thereby avoid the assessment of one distribution.

We can use partitions to assess probability distributions for features that are dependent on other features. For example, as is indicated in Figures 1 and 2, the probability distribution for FCB (prominent fibrocollagenous bands) depends on the degree of capsule thickening. To assess the probability distributions for FCB, we build a partition for every instance of CAP THICKENING. Figure 5 contains partitions for FCB, given two of the four instances of CAP THICKENING.

Using partitions, we decreased the number of probabilities required by Pathfinder's belief network from 75,000 to 14,000. Furthermore, the time we spent constructing partitions was less than 10% of the time we spent assessing probabilities. This observation may seem

surprising, given that a partition must be constructed for each conditioning instance of every feature. Two factors, however, contributed to the efficiency of the approach. First, the task of composing a single partition is straightforward. Apparently, people find it easy to make judgments of subset independence without assessing the probabilities underlying such judgments. Second, partitions often are identical or related from one feature to another. For example, as shown in Figure 5, the partitions for FCB given two instances of CAP THICKENING are identical. In constructing partitions, we used this close relationship to avoid constructing each partition from scratch.

5 SimNet: A Graphical Knowledge-Acquisition Tool

To construct Pathfinder, we created SimNet, an implementation of the belief-network, similarity-network, and partition representations on the Macintosh computer [5]. The figures shown in this article were created with SimNet.

In practice, an expert first uses SimNet to create a similarity graph. The expert then selects an edge of interest, and the program automatically sets up a belief-network template (containing only the disease node) from which the expert can construct the local belief network corresponding to the selected edge. As the local belief networks are created by the expert, SimNet automatically constructs the global belief network. The expert then uses partitions to assess the probability distributions associated with each feature in the global belief network.

6 An Evaluation of Belief Networks, Similarity Networks, and Partitions

The belief-network, similarity-network, and partition representations simplify the construction of normative expert systems. It is important, however, to ask whether or not the benefits of constructing an accurate diagnostic system using these representations outweigh the costs of constructing such a system. In a formal evaluation, we addressed this question [23].

In this evaluation, we examined two versions of Pathfinder: Pathfinder 2, the version of Pathfinder described in this article, and Pathfinder 1, an older version of Pathfinder, in which we made the assumption that all features are conditionally independent, given each disease.[4] We spent approximately 45 hours constructing Pathfinder 1; we did not use nor require a belief network, a similarity network, or partitions to build this version. In contrast, we spent approximately 80 hours constructing Pathfinder 2.

To assess the benefits of Pathfinder 2 relative to those of Pathfinder 1, we developed a procedure based in decision theory that measures the expected utility of a patient who receives a diagnosis from an expert system. Applying this procedure to a series of 53 randomly selected cases referred to the second author for diagnosis, we found that the increase in expected utility of a patient who receives a diagnosis from Pathfinder 2 over that of a patient who receives a diagnosis from Pathfinder 1 averaged $6000 per case. Thus, assuming

[4]Pathfinder II, described in the companion to this article, consists of Pathfinder 2 and the Intellipath delivery platform.

a combined hourly rate of $400 for the coauthors of this article, the additional effort to build Pathfinder 2 would more than pay for itself after only three cases had been run.

7 Summary

Probability-based representations can be practical tools for encoding and reasoning with uncertain medical knowledge. As we have seen, similarity networks and partitions have provided a cost-effective approach for the construction of the Pathfinder belief network. In addition, the Pathfinder belief network has provided efficient means for the computation of differential diagnoses. We hope that our discussion will encourage investigators to develop belief-network inference algorithms and extensions to the representation that will simplify further the construction and use of normative expert systems.

Acknowledgments

We thank Eric Horvitz, Keung-Chi Ng, Greg Cooper, and Lyn Dupré for reviewing earlier drafts of this manuscript. We also thank Henri Suermondt, Mark Fischinger, Marty Chavez, and especially Keung-Chi Ng for their assistance with programming and data management.

This research has been supported by the National Cancer Institute under Grant RO1CA51729-01A, and by the National Library of Medicine under Grant RO1LM04529.

References

[1] R.A. Howard and J.E. Matheson. Influence diagrams. In R.A. Howard and J.E. Matheson, editors, *Readings on the Principles and Applications of Decision Analysis*, volume II, pages 721–762. Strategic Decisions Group, Menlo Park, CA, 1981.

[2] J. Pearl. Reverend Bayes on inference engines: A distributed hierarchical approach. In *Proceedings AAAI-82 Second National Conference on Artificial Intelligence,* Pittsburgh, PA, pages 133–136. AAAI Press, Menlo Park, CA, August 1982.

[3] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Pathfinder research directions. Technical Report KSL-89-64, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, October 1985.

[4] D.E. Heckerman, E.J. Horvitz, and B.N. Nathwani. Update on the Pathfinder project. In *Proceedings of the Thirteenth Symposium on Computer Applications in Medical Care,* Washington, DC, pages 203–207. IEEE Computer Society Press, Silver Spring, MD, November 1989.

[5] D.E. Heckerman. *Probabilistic Similarity Networks*. MIT Press, Cambridge, MA, 1991.

[6] I.J. Good. A causal calculus (I). *British Journal of Philosophy of Science*, 11:305–318, 1961. Also in I.J. Good, *Good Thinking: The Foundations of Probability and Its Applications,* pages 197–217. University of Minnesota Press, Minneapolis, MN, 1983.

[7] I.J. Good. A causal calculus (II). *British Journal of Philosophy of Science*, 12:43–51, 1961. Also in I.J. Good, *Good Thinking: The Foundations of Probability and Its Applications,* pages 197–217. University of Minnesota Press, Minneapolis, MN, 1983.

[8] W.F. Rousseau. A method for computing probabilities in complex situations. Technical Report 6252-2, Center for Systems Research, Stanford University, Stanford, CA, May 1968.

[9] J. Pearl. *Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference*. Morgan Kaufmann, San Mateo, CA, 1988.

[10] T. Verma and J. Pearl. Causal networks: Semantics and expressiveness. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence,* Minneapolis, MN, pages 352–359. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1988. Also in Shachter, R., Levitt, T., Kanal, L., and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 4,* pages 69–76. North-Holland, New York, 1990.

[11] D. Geiger and J. Pearl. On the logic of causal models. In *Proceedings of the Fourth Workshop on Uncertainty in Artificial Intelligence,* Minneapolis, MN, pages 136–147. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1988. Also in Shachter, R., Levitt, T., Kanal, L., and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 4,* pages 3–14. North-Holland, New York, 1990.

[12] S.L. Lauritzen and D.J. Spiegelhalter. Local computations with probabilities on graphical structures and their application to expert systems. *J. Royal Statistical Society B*, 50:157–224, 1988.

[13] G.F. Cooper. *NESTOR: A Computer-based Medical Diagnostic Aid that Integrates Causal and Probabilistic Knowledge*. PhD thesis, Medical Computer Science Group, Stanford University, Stanford, CA, November 1984. Report HPP-84-48.

[14] R.A. Howard. Knowledge maps. *Management Science*, 35:903–922, 1989.

[15] R.D. Shachter. Evaluating influence diagrams. *Operations Research*, 34:871–882, 1986.

[16] J. Pearl. Fusion, propagation, and structuring in belief networks. *Artificial Intelligence*, 29:241–288, 1986.

[17] M. Henrion. Propagation of uncertainty by probabilistic logic sampling in Bayes' networks. In *Proceedings of the Second Workshop on Uncertainty in Artificial Intelligence,* Philadelphia, PA. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, August 1986. Also in Kanal, L. and Lemmer, J., editors, *Uncertainty in Artificial Intelligence 2,* pages 149–164. North-Holland, New York, 1988.

[18] G.F. Cooper. Bayesian belief-network inference using recursive decomposition. Technical Report KSL-90-05, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, January 1990.

[19] D. Geiger, T. Verma, and J. Pearl. Identifying independence in Bayesian networks. *Networks*, 20:507–534, 1990.

[20] H.J. Suermondt, G.F. Cooper, and D.E. Heckerman. A combination of cutset conditioning with clique-tree propagation in the pathfinder system. In *Proceedings of the Sixth Conference on Uncertainty in Artificial Intelligence,* Boston, MA, pages 273–279. Association for Uncertainty in Artificial Intelligence, Mountain View, CA, July 1990.

[21] D.E. Heckerman. Formalizing heuristic methods for reasoning with uncertainty. Technical Report KSL-88-07, Medical Computer Science Group, Section on Medical Informatics, Stanford University, Stanford, CA, May 1987.

[22] D.E. Heckerman. Probabilistic similarity networks. *Networks*, 20:607–636, 1990.

[23] D. Heckerman and B. Nathwani. An evaluation of the diagnostic accuracy of Pathfinder. *Computers and Biomedical Research*, In press.

Appendix: Glossary of Terms

Diseases of the Lymph Node

AIDS EARLY: AIDS, early phase

AIDS INVOLUTIONARY: AIDS, involutionary phase

AILD: Angio-immunoblastic lymphadenopathy

ALIP: Atypical lymphoplasmacytic and immunoblastic proliferation

AML: Acute myeloid leukemia

B-IMMUNOBLASTIC: Immunoblastic plasmacytoid diffuse lymphoma

CARCINOMA: Carcinoma

CAT SCRATCH DISEASE: Cat-scratch disease

CELLULAR PHASE NSHD: Cellular phase of nodular sclerosing Hodgkin's disease

DERMATOPATHIC LADEN: Dermatopathic lymphadenitis

DIFFUSE FIBROSIS HD: Diffuse fibrosis Hodgkin's disease

EM PLASMACYTOMA: Extramedullary plasmacytoma

FLORID FOLLIC HYPERP: Florid reactive follicular hyperperplasia

GLH HYALINE VACULAR: Giant lymph-node hyperplasia, hyaline vacular type

GLH PLASMA CELL TYPE: Giant lymph-node hyperplasia, plasma-cell type

GRANULOMATOUS LADEN: Granulomatous lymphadenitis

HAIRY CELL LEUKEMIA: Hairy cell leukemia

HISTIOCYTOSIS X: Histiocytosis x

IBL-LIKE T-CELL LYM: Immunoblastic lymphadenopathy-like T-cell lymphoma

INFECTIOUS MONO: Infectious mononucleosis

INTERFOLLICULAR HD: Interfollicular Hodgkin's disease

JAPANESE ATL: Japanese adult T-cell lymphoma

KAPOSIS SARCOMA: Kaposis sarcoma

L&H DIFFUSE HD: Lymphocytic and histiocytic diffuse Hodgkin's disease

L&H NODULAR HD: Lymphocytic and histiocytic nodular Hodgkin's disease

LARGE CELL, DIF: Large cell diffuse lymphoma

LARGE CELL, FOL: Large cell follicular lymphoma

LEPROSY: Leprosy

LYMPHANGIOGRAPHIC: Lymphangiography effect

LYMPHOBLASTIC: Lymphoblastic lymphoma

MALIG HISTIOCYTOSIS: Malignant histiocytosis

MANTLE ZONE: Mantle-zone lymphoma

MANTLE ZONE HYPERL: Mantle-zone hyperplasia

MAST CELL DISEASE: Mast-cell disease

MELANOMA: Melanoma

MIXED CELLULARITY HD: Mixed-cellularity Hodgkin's disease

MIXED, FCC DIF: Mixed (follicular center cell type) diffuse lymphoma

MIXED, FOL: Mixed (follicular center cell type) follicular lymphoma

MULTIPLE MYELOMA: Multiple myeloma

MYCOSIS FUNGOIDES: Mycosis fungoides

NECROTIZ NONKIKUCHI: NonKikuchi's necrotizing lymphadenitis

NECROTIZING KIKUCHI: Kikuchi's necrotizing lymphadenitis

NODULAR SCLEROSIS HD: Nodular sclerosing Hodgkin's disease

PLASMACYTOID LYCTIC: Small lymphocytic diffuse lymphoma with plasmacytoid features

RETICULAR TYPE HD: Reticular type Hodgkin's disease

RHEUMATOID ARTHRITIS: Rheumatoid arthritis

SARCOIDOSIS: Sarcoidosis

SHML: Sinus histiocytosis with massive lymphadenopathy

SINUS HYPERPLASIA: Sinus hyperplasia

SMALL CLEAVED, DIF: Small cleaved diffuse lymphoma

SMALL CLEAVED, FOL: Small cleaved follicular lymphoma

SMALL LYMPHOCYTIC: Small lymphocytic lymphoma

SMALL NONCLEAVED DIF: Small noncleaved diffuse lymphoma

SMALL NONCLEAVED FOL: Small noncleaved follicular lymphoma

SYNCYTIAL NSHD: Syncytial nodular sclerosing Hodgkin's disease

SYPHILIS: Syphilis

T-IMMUNOB LRG: Peripheral T-cell lymphoma, large-cell type

T-IMMUNOB MIX: Peripheral T-cell lymphoma, mixed-cell type

TOXOPLASMOSIS: Toxoplasmosis

TRUE HISTIOCYTIC: True histiocytic lymphoma

TUBERCULOSIS: Tuberculosis

VIRAL NOS: Viral lymphadenitis, not otherwise specified

WHIPPLE'S DISEASE: Whipple's disease

Features of the Lymph Node

ABR T-CELL PHENO: Abberrant T-cell phenotype in medium-sized or large lymphoid cells

ACID FAST STAIN: Acid fast stain

B GENE REARRANGEMENT: Immunoglobulin gene rearrangement

BNG HIST: Benign histiocytes not otherwise specified in the nonfollicular areas

BNG HIST FOAMY: Foamy benign histiocytes in the nonfollicular areas that do not contribute to mottling

BNG HIST LANGERHANS: Langerhans benign histiocytes in the nonfollicular areas

BNG HIST SS: Starry-sky benign histiocytes in the nonfollicular areas

CAP THICKENING: Capsule thickening (number of lymphocytes thick)

CARCINOMA CELLS: Carcinoma cells

CLASSIC SR: Classic Sternberg–Reed cells (number per 4-square-centimeter section)

DIL VASC SP: Vascular spaces dilated by red blood cells

EMPERIPOLESIS: Number of histiocytes showing emperipolesis

EOSIN MICROAB: Eosinophil microabscessess

EOSIN MYELO&META: Eosinophilic myelocytes and metamyelocytes

EOSINOPHILS: Eosinophils (not in microabcesses)

EPI HIST CLUS: Epithelioid histiocyte clusters

EPI HIST CLUS FOL EN: Epitheliod histiocyte clusters encroaching and/or within follicles

EPI HIST NONCLUSTERS: Epitheliod histiocyte nonclusters (percent of total cell population)

EXTRAVASC CLUS CLR C: Extravascular clusters of clear lymphoid cells

F % AREA: Percent area occupied by follicles

F CC CYTOLOGY: Cytology of follicular center cells in most follicles

F CENTERS ATROPHIC: Atrophic centers in any follicles

F CYTOLOGY COMP: Similar cells inside and outside of most follicles

F DEFINITION: Definition of follicles

F DENSITY: Follicle density

F HEMORRHAGES: Hemmorrhages in any of the follicles

F LYMPH INFIL: Lymphocyte infiltration of any follicles

F MANTLE ZONES: Follicle mantle zones in any follicles

F MIT FIGURES: Follicle mitotic figures in 10 high-power fields

F MZ CONCENTRIC RIMS: Mantle zone concentric rims in any follicles

F MZ STATUS: Follicle mantle zones

F POLARITY: Prominent polarity in any follicle

F RADIALLY PEN BV: Number of follicles showing radially penetrating blood vessels

F SS PATTERN: Follicle starry-sky histiocytes (average number in one 10X objective power)

FCB: Fibrocollagenous bands or sclerosis

FCB NODULES: Nodules formed by fibrocollagenous bands

FIBROSIS: Prominent fibrosis

FITE STAIN: Fite stain

FOLLICLES: Follicles

FOREIGN BODY: Foreign body (number in 4-square-centimeter section)

HAIRY CELLS: Hairy cells

HTLV I: HTLV I antibody test

HTLV III: HTLV III antibody test

INTRAVASC CLUS LYMPH: Intravascular clusters of lymphoid cells

KARYORRHEXIS: Karyorrhexis

L&H NODULES: Lymphocytic and hitiocytic nodules

L&H SR: Lymphocytic and hitiocytic variants of Sternberg–Reed cells (number in 4-square-centimeter section)

LACUNAR SR: Lacunar variants of Sternberg–Reed cells (number in 4-square-centimeter section)

LANGHANS: Langhans cells (number in 4-square-centimeter section)

LC LYSOZYME: Lysozyme positivity in medium-sized and/or large lymphoid cells

LEUKEMIC CELLS: Leukemic cells

LLC CHROMATIN: Chromatin of most large lymphoid cells

LLC CYTOPLASM: Cytoplasm of most large lymphoid cells

LLC EV CLUS: Large lymphoid cells in extravascular clusters of clear cells

LLC IDENTITY: Identity of most large lymphoid cells

LLC IV CLUS: Large lymphoid cells in intravascular clusters

LLC NUC SHP: Nuclear shape of most large lymphoid cells

LLC NUCLEOLI: Nucleolar features of most large lymphoid cells

LLC NUM: Number of large lymphoid cells in the nonfollicular areas (percent of total cell population)

LLC+MLC > 50%: Number of medium-sized and large lymphoid cells in the nonfollicular areas exceeds 50 percent of total cell population

LRG LMPH CELLS: Large lymphoid cells

MAST CELLS: Mast cells (number in 4-square-centimeter section)

MED LYMPH CELLS: Medium-sized lymphoid cells

MELANOMA CELLS: Melanoma cells

MITOTIC FIG: Mitotic figures in 10 high-power fields (nonfollicular areas)

MLC CHROMATIN: Chromatin structure of most medium-sized lymphoid cells

MLC CYTOPLASM: Cytoplasm of most medium-sized lymphoid cells

MLC EV CLUS: Medium-sized lymphoid cells in extravascular clusters of clear cells

MLC IV CLUS: Medium-sized lymphoid cells in intravascular clusters

MLC NUC SHP: Nuclear shape of most medium-sized lymphoid cells

MLC NUCLEOLI: Nucleolar features of most medium-sized lymphoid cells

MLC NUM: Number of Medium-sized lymphoid cells in the nonfollicular areas (percent of total cell population)

MONOCYT: Monocytoid cells (percent of total cell population)

MONONUCLEAR SR: Mononuclear variants of Sternberg–Reed cells (number in 4-square-centimeter section)

MOTTLING HIST: Mottling by langerhans or other histiocytes

MOTTLING LLC: Mottling by large lymphoid cells

MUMMY: Large mummified cells (number in 4-square-centimeter section)

NECROSIS: Necrosis

NEUTROPHIL MICROABSC: Neutrophil microabcessess

NEUTROPHILS: Neutrophils (not in microabcesses)

NONSIN NONFOL AREAS: Nonsinus nonfollicular areas

PAS STAIN: Strong PAS positivity in the histiocytes

PERICAP INFILTR: Pericapsular infiltration

PLASMA: Plasma cells in the nonfollicular areas (percent of total cell population)

PLASMA TYPE: Plasma cell type

PLEOMORPHIC SR: Pleomorphic variants of Sternberg–Reed cells (number in 4-square-centimeter section)

PSEUDOFOLLICLES: Pseudofollicles

PTGC: Progressively transformed germinal centers

RUSSELL&DUTCHER: Russell and/or Dutcher bodies

SARCOMA CELLS: Sarcoma cells

SCHAUMAN: Schauman cells

SIGNET-RING: Signet-ring cells

SINUSES: Sinuses

SLC CHROMATIN: Chromatin structure of most small lymphoid cells

SLC CYTOPLASM: Cytoplasm of most small lymphoid cells

SLC EV CLUS: Small lymphoid cells in extravascular clusters of clear cells

SLC IV CLUS: Small lymphoid cells in intravascular clusters

SLC NUC SHP: Nuclear shape of most small lymphoid cells

SLC NUM: Number of small lymphoid cells in the nonfollicular areas (percent of total cell population)

SML LYMPH CELLS: Small lymphoid cells

SR-LIKE: Sternberg–Reed-like cells (number in 4-square-centimeter section)

SYSTEMIC AIDS: Systemic AIDS

T GENE REARRANGEMENT: T-cell receptor gene rearrangement

TRANSITION FORMS: Transition forms (lymphoid cells having sizes other than the sizes of small, medium-sized, or large cells) in the nonfollicular areas

VASC CHANGES: Endarteritis or periarteritis

VASC PROLIF NONSLIT: Vascular proliferation (nonslitlike)

VASC PROLIF SLIT: Vascular proliferation slitlike