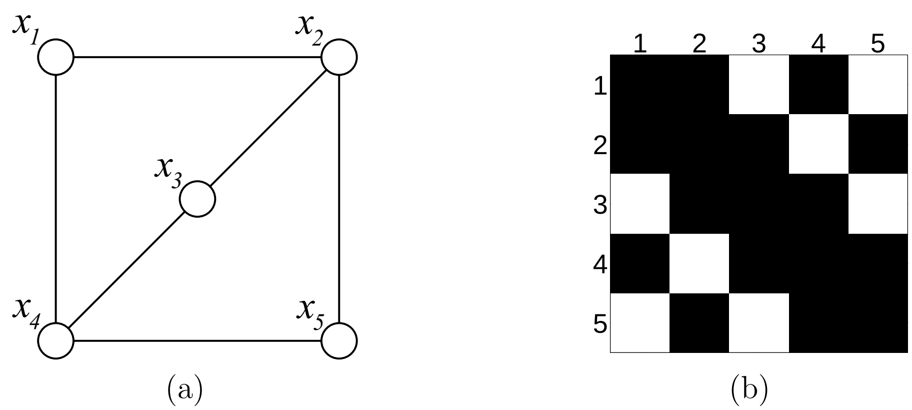

Figure 2-2: (a) Graphical model representing five jointly Gaussian random vectors. (b) Structure of the corresponding inverse covariance matrix P^{-1}, where black squares denote nonzero entries.

If $\mathcal{C}$ is the set of all cliques of $\mathcal{G}$, then a strictly positive distribution $p(x)$ is Markov with respect to $\mathcal{G}$ if and only if it can be written in the factorized form

$$p(x) = \frac{1}{Z} \prod_{C \in \mathcal{C}} \psi_C(x_C) \tag{2.13}$$

where $\psi_C(x_C)$ is an arbitrary positive function, or clique potential, defined over the elements of the clique C. Z is a normalization constant, sometimes called the partition function. For example, applying the Hammersley–Clifford Theorem to the graph in Figure 2-1(a), we find that the joint distribution $p(x)$ must factorize as

$$p(x_1, x_2, x_3, x_4, x_5) = \frac{1}{Z} \psi_{2,3}(x_2, x_3)\, \psi_{1,3,4}(x_1, x_3, x_4)\, \psi_{1,3,5}(x_1, x_3, x_5)$$

There are a variety of proofs of the Hammersley–Clifford Theorem [13, 18, 48]; see Clifford [21] for an interesting historical survey.

In many cases, the prior model is most naturally specified by "pairwise" clique potentials involving pairs of nodes which are connected by edges:

$$p(x) = \frac{1}{Z} \prod_{(s,t) \in \mathcal{E}} \psi_{s,t}(x_s, x_t) \tag{2.14}$$

An arbitrary Markov random field may be represented using only pairwise clique potentials by appropriately clustering nodes in the original graph [84]. However,