Combining Proposition 2.1 with the Markov properties implied by equations (2.8, 2.9), we may show that the inverse covariance matrix has the following structure [48, 69]:

Theorem 2.2. Let $x \sim \mathcal{N}(0, P)$ be a Gaussian stochastic process which is Markov with respect to an undirected graph $\mathcal{G} = (\mathcal{V}, \mathcal{E})$. Assume that x is *not* Markov with respect to any $\mathcal{G}' = (\mathcal{V}, \mathcal{E}')$ such that $\mathcal{E}' \subsetneq \mathcal{E}$, and partition $J = P^{-1}$ into a $|\mathcal{V}| \times |\mathcal{V}|$ grid according to the dimensions of the node variables. Then for any $s, t \in \mathcal{V}$ such that $s \neq t$, $J_{s,t} = J_{t,s}^T$ will be nonzero if and only if $(s, t) \in \mathcal{E}$.

Proof. Consider any $s, t \in \mathcal{V}$ such that $s \neq t$. Combining equation (2.11) with basic properties of the singular value decomposition, we see that $J_{s,t}$ will be nonzero if and only if $\operatorname{cov}\left(x_s, x_t \,\middle|\, x_{N(s,t)}\right) \neq \mathbf{0}$. Since x_s and x_t are jointly Gaussian, it follows that $J_{s,t}$ will be zero if and only if x_s and x_t are conditionally independent given $x_{N(s,t)}$.

Suppose that $(s, t) \notin \mathcal{E}$, then $N(s, t)$ separates nodes s and t, and by the Markov properties of $\mathcal{G}$, x_s and x_t must be conditionally independent. This in turn implies $J_{s,t} = \mathbf{0}$. Alternatively, if $(s, t) \in \mathcal{E}$, then $N(s, t)$ does not separate s and t. By assumption, x is not Markov with respect to the subgraph created by removing edge (s, t), so x_s and x_t must be conditionally dependent, and therefore $J_{s,t} \neq \mathbf{0}$. □

Figure 2-2 illustrates Theorem 2.2 for a small sample graph. In most graphical models, each node is only connected to a small subset of the other nodes. Theorem 2.2 then shows that P^{-1} will be a *sparse* matrix with a small (relative to N) number of nonzero entries in each row and column. This sparsity is the fundamental reason for the existence of the efficient inference algorithms discussed in §2.3.

2.2.3 Parameterization of Gaussian Markov Random Fields

The sparse structure exhibited by the inverse covariance matrix of a Gaussian Markov random field is a manifestation of the constraints which the Markov properties discussed in §2.2.1 place on $p(x)$. Similar constraints also hold for more general undirected graphical models. In particular, the Hammersley–Clifford Theorem relates the Markov properties implied by $\mathcal{G}$ to a factorization of the probability distribution $p(x)$.