Proofs

In order to prove Proposition 3, it is convenient to first prove the following lemma.

Lemma A. Consider a connected GMRF $G = (V, E; J)$ parameterized by precision matrix $J$ and a unique path $P$ embedded in $G$. The marginal precision matrix $J_P$ has block off-diagonal elements identical to those in the submatrix of $J$ corresponding to variables in $P$, and block diagonal elements that are the Schur complements of the submatrices corresponding to the sidegraphs separated by $P$.

Proof. Assume without loss of generality that the unique path under consideration is $(1, 2, \ldots, k)$. Because $P_{1:k}$ is unique, the graph $\tilde{G}$ induced by $V \setminus P_{1:k}$ can be thought of as the union of conditionally independent “sidegraphs” $\tilde{G}_1, \ldots, \tilde{G}_k$, each of which is connected in $\tilde{G}$ to a single node in $P_{1:k}$ (see Figure 2a). Let $J_{P_{1:k}}$ denote the (block tridiagonal) matrix parameterizing the joint potential (i.e., the product of singleton and edge potentials in the factorization of the full joint distribution of $x$) over the chain $(1, \ldots, k)$. For all $i \in \{1, \ldots, k\}$, let $J_{\tilde{G}_i}$ be the matrix parameterizing the potentials over edges between $i$ and $N(i) \setminus P_{1:k}$. Likewise, let $J_{\tilde{G}_i}$ denote the matrix parameterizing the joint potential over the subgraph $\tilde{G}_i$.

Now consider a permutation to the last $n - k$ components of $J$ such that $J_{\tilde{G}_1}, \ldots, J_{\tilde{G}_k}$ are ordered as such, whereby

$$J \triangleq \begin{bmatrix} J_{P_{1:k}} & J_{P_{1:k}, \tilde{G}} \\ J_{P_{1:k}, \tilde{G}}^T & J_{\tilde{G}} \end{bmatrix}. $$

In this permuted matrix, $J_{\tilde{G}}$ is block diagonal – due to conditional independence of the sidegraphs – with elements $J_{\tilde{G}_i}$. Similarly, the upper-right block submatrix $J_{P_{1:k}, \tilde{G}}$ is also block diagonal with elements $J_{i, \tilde{G}_i}$. Thus, the marginal distribution $p_{x_1, \ldots, x_k}$ is parameterized by a precision matrix

$$J'_{P_{1:k}} = J_{P_{1:k}} - J_{P_{1:k}, \tilde{G}} J_{\tilde{G}}^{-1} J_{\tilde{G}, P_{1:k}};$$

where the subtractive term is a product of block diagonal matrices and, thus, is itself a block diagonal matrix. Therefore, the marginal precision matrix $J'_{P_{1:k}}$ has block off-diagonal elements identical to those of the submatrix $J_{P_{1:k}}$ of the (full) joint precision matrix; each block diagonal element is the Schur complement of each $J_{\tilde{G}_i}$, $i = 1, \ldots, k$.

Remark B. Lemma A implies that if Proposition 3 holds for any chain of length $k$ between nodes $u$ and $v$, it must also hold for the more general class of graphs in which $|P(u, v)| = 1$ (i.e., there is a unique path between $u$ and $v$, but there are sidegraphs attached to each vertex in the path). Therefore, we need only prove Proposition 3 for chains of arbitrary length. Furthermore, conditioning only severs nodes from the graph component considered; provided $C$ is not a separator for $u, v$, in which case $I(x_u; x_v | x_C) = 0$, we need only prove Proposition 3 for the case where $C = \emptyset$. 
Proof of Proposition 3. We proceed by induction on the length $k$ of the chain. The base case considered is a chain of length $k = 3$, for which the precision matrix is

$$
\begin{bmatrix}
J_{11} & J_{12} & 0 \\
J_{12} & J_{22} & J_{23} \\
0 & J_{23} & J_{33}
\end{bmatrix}.
$$

By Remark 2, we need only show that $|\rho_{13}| = |\rho_{12}| \cdot |\rho_{23}|$. We have

$$
J = \frac{1}{\det(J)} \begin{bmatrix}
J_{22}J_{33} - J_{23}^2 & -J_{12}J_{33} & J_{12}J_{23} \\
-J_{12}J_{33} & J_{11}J_{33} - J_{13}^2 & -J_{11}J_{23} \\
J_{12}J_{23} & -J_{11}J_{23} & J_{11}J_{22} - J_{12}^2
\end{bmatrix},
$$

from which it is straightforward to confirm that

$$
\rho_{12} = -J_{12}J_{33}/\sqrt{J_{11}J_{33}(J_{22}J_{33} - J_{23}^2)}
$$

$$
\rho_{23} = -J_{11}J_{22}/\sqrt{J_{11}J_{33}(J_{11}J_{22} - J_{12}^2)}
$$

$$
\rho_{13} = J_{12}J_{23}/\sqrt{(J_{11}J_{22} - J_{12}^2)(J_{22}J_{33} - J_{23}^2)}
$$

$$
= \rho_{12} \cdot \rho_{23},
$$

thus proving the base case.

Now, assume the result of the proposition holds for a unique path $\tilde{P}_{1:k}$ of length $k$ embedded in graph $G$, and consider node $k + 1 \in N(k) \setminus \tilde{P}_{1:k}$. By Lemma A, we can restrict our attention to the marginal chain over $(1, \ldots, k, k + 1)$. Pairwise decomposition on subchains of length $k$ yields an expression for $I(x_1:x_k)$, which by (4) can alternatively be expressed in terms of the determinants of marginal precision matrices. Therefore, if one marginalizes nodes in $\{2, \ldots, k - 1\}$ (under any elimination ordering), one is left with a graph over nodes $1$, $k$, and $k + 1$. The MI between $k$ and $k + 1$, which are adjacent in $G$, can be computed from the local GaBP messages comprising the marginal node and edge precision matrices. On this 3-node network, for which the base case holds,

$$
W(1; k + 1) = W(k; k + 1) + W(1; k)
$$

$$
= W(k; k + 1) + \sum_{\{i,j\} \in E_{1,k}} W(i; j)
$$

$$
= \sum_{\{i,j\} \in E_{1,k+1}} W(i; j).
$$

Therefore, Proposition 3 holds for chains of arbitrary length and, by Lemma A, unique paths of arbitrary length in GMRFs with scalar variables. \hfill \Box

Proof of Proposition 6. The proof follows closely that of Proposition 3. By Lemma A, assume without loss of generality that the unique path under consideration is a vectorial chain of length $k$ with sequentially indexed nodes, i.e., $(1, 2, \ldots, k - 1, k)$. Thinness of edges $\{i, j\} \in E$ implies $W(i; j) = \log |\rho_{ij}|$, as before. Let $i \in \{2, \ldots, k - 1\}$ be an arbitrary interstitial node. On section $(i - 1, i, i + 1)$ of the chain, thinness of $\{i - 1, i\}$ and $\{i, i + 1\}$ implies that on $\tilde{P}_{i-1:k}$, there exists one inlet to and one outlet from $i$. Let $m$ and $q$ denote the column of $J_{i-1,i}$ and row of $J_{i,i+1}$, respectively, containing nonzero elements. Then the path through the internal structure of node $i$ can be simplified by marginalizing out, as computed via Schur complement from $J'_{i}$ in $O(d^2)$, all scalar elements of $i$ except $m$ and $q$. Thus, $\zeta_i(u, v)$ is merely the warped mutual information between $m$ and $q$, and problem reduces to a scalar chain with alternating $W$ and $\zeta$ terms. \hfill \Box