Factoring a Graph in Polynomial Time

PETER WINKLER

We present a polynomial-time algorithm for deciding whether a given connected graph is a non-trivial Cartesian product. The method entails first representing the graph as an isometric subgraph of a Cartesian product of graphs, then finding a suitable partition of the factors.

The Cartesian product $G \times H$ of graphs $G$ and $H$ has as vertices the pairs $(g, h)$ with $g$ a vertex of $G$ and $h$ a vertex of $H$; $(g_1, h_1)$ is connected by an edge to $(g_2, h_2)$ in $G \times H$ just when $\{g_1, g_2\}$ is an edge of $G$ and $h_1 = h_2$, or when $g_1 = g_2$ and $\{h_1, h_2\}$ is an edge of $H$. The Cartesian product admits unique factorization (Sabidussi [4]) but until recently no efficient algorithm was known for producing such a factorization.

If unconnected graphs are permitted then the factorization problem is at least as difficult as ‘graph isomorphism’; for, one could determine whether two connected graphs $G$ and $H$ are isomorphic by deciding whether a graph with two vertices and no edge is a factor of the disjoint union of $G$ and $H$. The question of whether there is a polynomial algorithm for deciding if a connected graph is a non-trivial Cartesian product—equivalently, for finding its unique factorization—was posed by Welsh [5], Imrich [3], and probably Sabidussi as well. Recently, this question was settled in the affirmative (independently of this author) by Feigenbaum, Hershberger and Schäffer [1] using towers of equivalence relations. Our methods are completely different, making use of results in [2] in which graphs are regarded as metric spaces.

**THEOREM.** There is a polynomial algorithm for deciding whether a given connected graph $G$ is a non-trivial Cartesian product, and for finding the prime factorization of $G$.

**PROOF.** If $u$ and $v$ are two vertices of $G$, let $d(u, v)$ be the number of edges in a shortest path between them. Define the relation $\beta$ between edges of the graph as follows: if $e = (x, y)$ and $f = (u, v)$ then $ef \beta$ if $d(x, u) + d(y, v) = d(x, v) + d(y, u)$. Let $E$ be the edge-set of $G$ and $E_1, \ldots, E_k$ the equivalence classes of the transitive closure of $\beta$. The factor graph $G_i$, $1 \leq i \leq k$, is formed from $G$ by contracting every edge not in $E_i$ to a point; concatenating the natural projections $f_i: G \to G_i$ yields a map $f$ from $G$ to the Cartesian product of the $G_i$s. This product may itself be exponential in size, but we do not need to list its vertices; in fact even the structure of the $G_i$s is irrelevant until later.

The first lemma lists some useful facts about this ‘canonical representation’ $f: G \to \prod G_i$; all follow immediately from Theorems 1–3 in [2].

**LEMMA 1.** The map $f$ is isometric, i.e. $d(u, v) = \Sigma_{i=1}^k d_{G_i}(f_i(u), f_i(v))$ for any pair $u, v$ of vertices in $G$; further, each $G_i$ has at least two vertices and the projections $f_i: G \to G_i$ are surjective. The number of factors $k$ is called the dimension of $G$ and $f$ is the unique embedding of $G$ with the above properties and at least $k$ factors; the dimension of each $G_i$ is 1.

If $G$ is the Cartesian product of graphs $H_1$ and $H_2$ then there is a partition of the index set $S = \{1, 2, \ldots, k\}$ into sets $S_1$ and $S_2$, and canonical representations $h_j: H_j \to \Pi_{i \in S_j} G_i$ for
\[ G = H_1 \times H_2 \]
\[ f \]
\[ h_1 \]
\[ h_2 \]
\[ \prod_{i \in S} G_i = \prod_{i \in S_1} G_i \times \prod_{i \in S_2} G_i \]

is a commuting diagram. Conversely, any partition of \( S \) into \( S_1 \) and \( S_2 \) induces maps \( h_i: G \to \Pi_{i \in S_i} G_i \); if \( H_i \) is the image of \( G \) under \( h_i \) then \( h_1 \times h_2 \) is an isometric embedding of \( G \) into \( H_1 \times H_2 \). If this map happens to be surjective, we have \( G = H_1 \times H_2 \); otherwise there can be no factoring of \( G \) corresponding to the partition \((S_1, S_2)\).

Our strategy now is as follows. We compute the graphs \( G_i \) and arbitrarily assign to each the vertex set \( M_j = \{1, 2, \ldots, m_j\} \). Then the canonical representation \( f \) assigns to each vertex \( v \) of \( G \) a sequence \((v_1, \ldots, v_k)\) of numbers, \( 1 \leq v_i \leq m_i \). Let \( U \) be \( \{(v_1, \ldots, v_k): v \text{ a vertex of } G\} \), so that \( U \subset \Pi_{i \in T} M_i \) and \( |U| = n \), the number of vertices of \( G \).

If \((S_1, S_2)\) is a partition of \( S = \{1, 2, \ldots, k\} \), let \( U_1 \) and \( U_2 \) be the images of \( U \) in \( \Pi_{i \in S_1} M_i \) and \( \Pi_{i \in S_2} M_i \), respectively. It follows from Lemma 1 that \((S_1, S_2)\) is a 'good' partition, i.e. induces a factoring of \( G \), precisely when \( U = U_1 \times U_2 \).

Unfortunately there may be exponentially many partitions so we cannot merely test each one for the above property. Instead, call a subset \( T \) of \( S \) complete if \( U_T = \Pi_{i \in T} M_i \), that is, if the projection of \( U \) into \( \Pi_{i \in T} M_i \) is surjective. We have:

**Lemma 2.** Every singleton \( \{i\} \subset S \) is complete, and every subset of a complete set is complete. If \( S \) itself is complete then \( G = \prod_{i=1}^k G_i \) and thus every partition of \( S \) is good; in particular, \( G \) is factorable iff \( k > 1 \).

Note that in fact \( k = 1 \) is the 'usual' case: most random graphs have dimension 1 and are thus quickly seen to be unfactorable. The dimension may be as much as \( n - 1 \) (in the case of a tree, which, as it happens, is also unfactorable.)

**Lemma 3.** A given subset \( T \) of \( S \) can be tested for completeness in polynomial time.

Of course \( \Pi_{i \in T} M_i \) may be exponentially large but if \( \Pi_{i \in T} m_i \) is larger than \( n \) then \( T \) cannot be complete.

We say that a subset \( T \) of \( S \) is **minimally incomplete** if it is not complete but every proper subset of it is complete.

**Lemma 4.** If \( S \) is not complete then a minimally incomplete subset of \( S \) can be found in polynomial time.

Every incomplete set contains a set which is minimally incomplete. We begin by checking all subsets of \( S \) of cardinality \( k - 1 \); if they are all complete then \( S \) itself is minimally incomplete. Otherwise some \( k - 1 \)-set \( T \) is incomplete and we test all of \( T \)'s \( k - 2 \)-subsets, etc. Altogether fewer than \( n^2 \) completeness tests are necessary.

Now comes the crux.

**Lemma 5.** Let \( T \) be a minimally incomplete subset of \( S \) and let \((S_1, S_2)\) be a good partition of \( S \). Then either \( T \subset S_1 \) or \( T \subset S_2 \).
Factoring a graph in polynomial time

If not then there is a partition \((T_1, T_2)\) of \(T\) with \(T_1 \subset S_1\) and \(T_2 \subset S_2\); since \(T_1\) and \(T_2\) are both complete, \(T\) itself is complete with respect to \(U_1 \times U_2\). Since \(T\) is incomplete with respect to \(U\), we cannot have \(U = U_1 \times U_2\) and a contradiction has been reached.

Having found a minimally incomplete subset \(T\), we may now replace the graphs \(\{G_i; i \in T\}\) by a new graph \(G_T\) having vertex set \(\{1, 2, \ldots, |U_T|\}\). Structurally, \(G_T\) is just the subgraph of \(\Pi_{i \in T} G_i\) induced by the image of \(G\) under the map \(\Pi_{i \in T} f_i\). Since \(G_T\) cannot be split by a factorization of \(G\), we are reduced to considering the new representation \(G \to G_T \times \Pi_{i \not\in T} G_i\) which has fewer than \(k\) factors (recall \(|T| \geq 2\) since singletons are complete).

We now repeat the process with the new representation, continuing until we have a representation whose index set is complete. This final representation (which may well be the identity map \(G \to G\)) is then precisely the unique prime factorization of \(G\) as a Cartesian product.

As an example consider the graph \(G\) with vertex set \(\{x_1, x_2, \ldots, x_{12}\}\) and the adjacency matrix given in Figure 1. After computing the distance matrix of \(G\) and the transitive closure of the relation \(\beta\) we find that \(k = 4\) and the canonical factor graphs \(G_1, \ldots, G_4\) are all single edges, i.e. isomorphic to \(K_2\). We will number the \(G_i\)s and their vertices in such a way that the edge \((x_1, x_4)\) lies in \(E_1\), \((x_1, x_8)\) in \(E_2\), \((x_1, x_{11})\) in \(E_3\) and \((x_2, x_{10})\) in \(E_4\); and \(f(x_1) = (1, 1, 1, 1)\). The listing for the set \(U\) given in Figure 2 is then obtained.

\[
\begin{array}{ccccccccc}
  x_1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
  x_2 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
  x_3 & 0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 \\
  x_4 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  x_5 & 0 & 0 & 1 & 1 & 0 & 1 & 0 & 0 & 0 \\
  x_6 & 0 & 0 & 0 & 1 & 0 & 1 & 0 & 0 & 1 \\
  x_7 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
  x_8 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
  x_9 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
  x_{10} & 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
  x_{11} & 1 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 0 \\
  x_{12} & 0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
\end{array}
\]

**Figure 1.**

\[
\begin{array}{cccc}
  1 & 2 & 3 & 4 \\
  x_1 & 1 & 1 & 1 \\
  x_2 & 2 & 2 & 1 \\
  x_3 & 1 & 1 & 2 \\
  x_4 & 2 & 1 & 1 \\
  x_5 & 1 & 2 & 2 \\
  x_6 & 1 & 2 & 1 \\
  x_7 & 2 & 1 & 2 \\
  x_8 & 1 & 2 & 1 \\
  x_9 & 2 & 2 & 2 \\
  x_{10} & 2 & 2 & 1 \\
  x_{11} & 1 & 1 & 2 \\
  x_{12} & 2 & 1 & 1 \\
\end{array}
\]

**Figure 2.**

Since \(2^4 > 12\), \(S = \{1, 2, 3, 4\}\) itself is incomplete. Its subsets \(\{1, 2, 3\}, \{1, 2, 4\}\) and \(\{2, 3, 4\}\) are complete but \(\{1, 3, 4\}\) is not: there is no vector in \(U\) of the form \((1, \cdot, 1, 2)\) or \((2, \cdot, 2, 1)\). The subsets \(\{1, 3\}, \{1, 4\},\) and \(\{3, 4\}\) are complete so \(T = \{1, 3, 4\}\) is minimally incomplete and can be contracted to a single factor. By replacing \((1, \cdot, 1, 1)\) by \((1, \cdot, 2, 1)\)
by 2, (1, ·, 2, 2) by 3, (2, ·, 1, 1) by 4, (2, ·, 1, 2) by 5 and (2, ·, 2, 2) by 6 one obtains the picture for $G_T$ given in Figure 3. The new representation of $G$ is given in Figure 4. Now the full index set \{2, $T$\} is complete so that $G = G_2 \times G_T \cong K_2 \times C_6$; the isomorphism can be read from the tables.

\begin{figure}[h]
\centering
\includegraphics[width=0.3\textwidth]{figure3.png}
\caption{Figure 3.}
\end{figure}

\begin{figure}[h]
\centering
\begin{tabular}{c|cc}
  & $2$ & $T$ \\
\hline
$x_1$ & 1 & 1 \\
$x_2$ & 2 & 5 \\
$x_3$ & 1 & 3 \\
$x_4$ & 1 & 4 \\
$x_5$ & 2 & 3 \\
$x_6$ & 2 & 2 \\
$x_7$ & 1 & 6 \\
$x_8$ & 2 & 1 \\
$x_9$ & 2 & 6 \\
$x_{10}$ & 2 & 4 \\
$x_{11}$ & 1 & 2 \\
$x_{12}$ & 1 & 5 \\
\end{tabular}
\caption{Figure 4.}
\end{figure}

ACKNOWLEDGMENT

Research supported by N.S.F. grant MCS 84-02054, O.N.R. grant NR539016/5-17-85410, and the Alexander von Humboldt Foundation.

REFERENCES


Received 26 March 1985

PETER WINKLER

Department of Mathematics and Computer Science,
Emory University,
Atlanta, GA 30322, U.S.A.