# Split Graphs

Abel Romer

November 6, 2020

**Abstract**

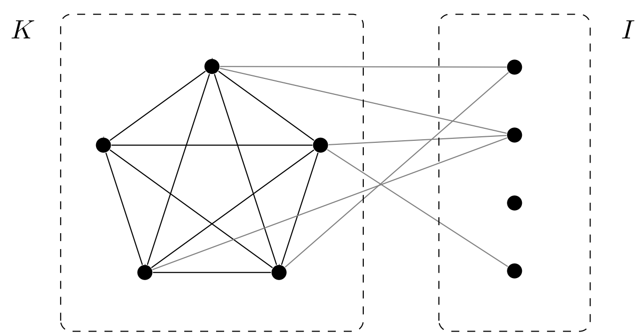
We introduce a class of graphs called *split graphs* and explore their attributes. We provide three methods for determining whether an arbitrary graph is split, and show that these characterizations are correct. Finally, we examine how split graphs provide computational advantage over arbitrary graphs in traditionally difficult optimization problems.

## 1 Introduction

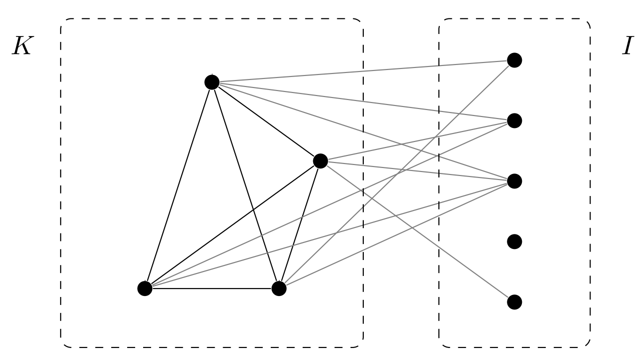
A simple graph is said to be *split* if its vertex set can be partitioned into a clique and an independent set. We formalize this idea as follows:

**Definition 1.1.** *A graph $G$ is split if its vertex set $V(G)$ can be partitioned into two disjoint sets $K$ and $I$ such that $K$ is a clique and $I$ is an independent set.*

Let's consider a simple example:




Here, $K$ is a $K_5$ and $I$ is a set of 4 independent vertices. We might make the observation that this is in fact not the only partition $G = (K, I)$ that forms a split graph. We could easily move any $v$ from $K$ to $I$, so long as for all $u \in I$, $uv \notin G$ (i.e. $v$ is not adjacent to any vertex in $I$). In our example, we can move the leftmost vertex in the clique $K$ to the independent set $I$, giving us the following split:



We will examine some more properties of split graphs below.

## 1.1 Basic properties

One simple observation we can make about split graphs is that they are *self-complementary* (i.e. if $G$ is a split graph, then $\overline{G}$ is a split graph). This follows immediately from the fact that the complement of a clique is an independent set. A slightly less obvious result claims that split graphs are a subset of *chordal graphs.*

**Definition 1.2.** *A chordal graph, also known as a triangulated graph, is a graph $G$ where every cycle $C$ of length greater than 4 has a chord (an edge not in $C$ that connects two vertices in $C$).*

**Theorem 1.1.** *Let $G$ be a graph. Then $G$ is split if and only if $G$ and $\overline{G}$ are chordal.*

*Proof:* $\Longrightarrow$ Suppose $G$ is split and contains a chordless cycle $C$ of length $\geq 4$. At least 1 vertex and at most 2 vertices of $C$ are in $K$. Since any two vertices in $K$ are connected, this implies that $I$ must contain at least 2 connected vertices. Therefore, $G$ must be chordal. If $G$ is split, then $\overline{G}$ is also split, and so $\overline{G}$ must also be chordal.

$\Longleftarrow$ This proof is due to Foldes and Hammer [1]. Suppose both $G$ and $\overline{G}$ chordal. It follows that the largest induced cycles in both $G$ and $\overline{G}$ are triangles. Let $K$ be the maximum clique in $G$. If there are multiple maximum cliques, let $K$ be the clique such that $e(G[V(G) \setminus K])$ (the edge set of the graph induced by $V(G) \setminus K$) is minimized. We show that $V(G) \setminus K$ is independent.

Suppose for contradiction that $x, y$ are two adjacent vertices in $V(G) \setminus K$. We claim that there exist two distinct vertices $u, v \in K$ such that $x \nsim u$ and $y \nsim v$ (i.e. $xu \notin E(G)$ and $yv \notin E(G)$). Clearly, neither $x$ nor $y$ is adjacent to all vertices in $K$, as this would imply that $K$ is not maximum. Additionally, if $x$ and $y$ were adjacent to all vertices in $K$ except some vertex $z$, then we could form a larger clique $K' = K \setminus \{z\} \cup \{x, y\}$. We also claim that exactly one of the edges in $\{xv, yu\}$ is in $E(G)$. If both $xv, yu \in E(G)$, then $(x, y, u, v)$ is an induced $C_4$ in $G$. If neither edge is in $E(G)$, then $(x, v, y, u)$ is an induced $C_4$ in $\overline{G}$. Without loss of generality, suppose $yu \in E(G)$.

We claim that $y$ is adjacent to all vertices of $K$ except $v$. Suppose otherwise. Then there exists some vertex $w \in K$, $w \neq v$ such that $yw \notin E(G)$. Either $x$ is adjacent to $w$ or $x$ is not adjacent to $w$. If $x \sim w$, then $(x, w, u, y)$ forms an induced $C_4$ in $G$. If $x \nsim w$, then $(x, v, y, w)$ is an induced $C_4$ in $\overline{G}$.

Let's recap. We have $x, y \in V(G) \setminus K$, and $u, v \in K$. We have shown that if $xy \in E(G \setminus K)$, then $y$ is adjacent to all vertices in $K$ except $v$, and $x$ is not adjacent to $u$ or $v$.

We show that if any other vertex $t$ in $V(G) \setminus K$ is not adjacent to $y$, then $t$ is not adjacent to $v$. Suppose otherwise and let $t$ be a vertex in $V(G) \setminus K$ adjacent to $v$ and not adjacent to $y$. Then $t$ must be adjacent to $x$, or otherwise $(x, t, y, v)$ would form an induced $C_4$ in $\overline{G}$. However, if $t \sim x$, then $(x, t, v, u, y)$ forms an induced $C_5$ in $G$.

We propose that there exists another clique $K'$ in $G$ that provides a smaller edge set in the graph induced by $V(G) \setminus K'$. Let this graph be $K' = K \setminus \{v\} \cup y$. Then the edge set of the graph induced by $V(G) \setminus K'$ contains at least one fewer edge than the graph induced by $V(G) \setminus K$ (namely, we lose the edge $xy$ and do not gain a corresponding edge $xv$). This is a contradiction, and we conclude that no two vertices in $V(G) \setminus K$ are adjacent.

This theorem provides us with a specific characterization of split graphs as *self-complementary chordal graphs.* From the fact that split graphs are a subset of chordal graphs and chordal graphs are perfect, we can make one final observation:

**Theorem 1.2.** *If $G$ is a split graph, then $G$ is perfect.*

As a final note:

**Definition 1.3.** *A graph $G$ is* perfect *if and only if for every induced subgraph $H$ of $G$, the chromatic number of $H$ is equal to the size of the maximum clique in $H$ (i.e. $\chi(H) = \omega(H)$).*

## 1.2 Motivating questions

Here are some additional questions to consider:

- How do we determine if a graph $G$ is split?
- What are the attributes of a split graph?
- What is the maximum independent set $I$? What is the maximum clique $K$? How do we find them?
- What is the minimum clique cover?
- What about colorings?

We will address these questions in roughly the order they appear.

# 2 Characterizations

## 2.1 Degree sequence and vertex ordering

How might we identify a split graph $G = (K, I)$? Since a clique is maximally connected and an independent set is minimally connected, it seems reasonable to suggest that whether or not a graph is split might have something to do with the degrees of its vertices. In fact, we can see that the number of edges between $K$ and $I$ is exactly the same as the total degree of all vertices in $I$ (i.e. $\sum_{v\in I} d(v)$). Suppose $|K| = k$. Then, since the total degree of $G[K]$ (the subgraph of $G$ induced by $K$) is $k(k-1)$, we should expect that

$$\sum_{v\in K} d(v) - \sum_{v\in I} d(v) = \sum_{v\in G[K]} d(v) = k(k-1).$$

Furthermore, if $K$ is maximum, we should expect that every vertex in $K$ has degree at least as large as every vertex in $I$. Otherwise, there would be some vertex in $I$ that is connected to all vertices in $K$, violating our assumption that $K$ was maximum. Therefore, we should be able to cut the degree sequence $(d_1, d_2, ...., d_n)$ of a split graph at some value $i$ into the sets $K$ and $I$. Together, these observations give us the following theorem:

**Theorem 2.1.** *Let $S = (v_1, v_2, ..., v_n)$ be an ordering of the vertex set $V(G)$ such that $v_i \geq v_{i+1}$ for all $v_i \in S$. Then, $G$ is a split graph if and only if there exists a value $m = 1, ..., n$ such that summing over $S$ gives:*

$$\sum_{i=1}^{m} d(v_i) - \sum_{i=m+1}^{n} d(v_i) = m(m-1).$$

*Proof:* $\Longrightarrow$ Suppose $G$ is split. Then $V(G)$ can be partitioned into sets $K$ and $I$. Let $K$ be a maximum clique. Let $m = n(K)$. We claim that $K = (v_1, v_2, ..., v_n)$ and $I = (v_{n+1}, ..., v_m)$. Suppose otherwise. Then there exists some $u \in K$, $v \in I$ such that $d(u) < d(v)$. However, if this is the case, then $K \cup v$ must be a clique that is strictly larger than $K$. This violates our assumption that $K$ is maximum. Therefore, $\sum_{i=1}^{m} d(v_i)$ is equal to the number of edges in $G[K]$ plus the edges between $K$ and $I$. Since the number of edges between $K$ and $I$ is given by $\sum_{i=m+1}^{n} d(v_i)$, we have:

$$\sum_{i=1}^{m} d(v_i) = \sum_{i=m+1}^{n} d(v_i) + m(m-1).$$

$\Longleftarrow$ The proof is presented in Hammer and Simeone (1981) [2].

There is another vertex ordering that may be used to characterize split graphs. In a 2018 paper [3], David Wood presents the following characterization:

**Theorem 2.2** (Wood)**.** *A graph G is split if and only if G has a vertex ordering* $(v_1, v_2, ..., v_n)$ *such that for all* $i < j < k$*:*

$$v_iv_j \in E(G) \Rightarrow v_jv_k \in E(G).$$

To understand the idea behind this theorem, note that such an ordering requires that the first $m$ vertices be independent, while the remaining $n - m$ vertices form a clique. Suppose otherwise. Let $v_i, v_j \notin K$ and $v_iv_j \in E(G)$. Then there must exist a $k$ such that $v_jv_k \notin E(G)$, because otherwise $v_j$ would be part of $K$. Therefore, $v_iv_j \notin E(G)$, and $(v_1, ..., v_m)$ is independent.

## 2.2 Forbidden subgraphs

Split graphs can also be characterized by their forbidden subgraphs. These are a set of induced subgraphs in $G$ whose absence (or presence) determines whether $G$ is (or is not) split. A result from Foldes and Hammer in 1977 [1] provides the following theorem:

**Theorem 2.3** (Foldes and Hammer)**.** *A graph G is split if and only if it does not contain a* $2K_2$*,* $C_4$*, or* $C_5$ *as an induced subgraph.*

*Proof:* $\Longrightarrow$ This follows almost directly from Theorem 1.1. Since $G$ is a split graph, both $G$ and $\overline{G}$ are chordal (i.e. the largest induced cycles in $G$ and $\overline{G}$ are triangles). Note that $2K_2$ is the complement of $C_4$. Therefore, $G$ cannot contain $2K_2$, $C_4$, or $C_5$ as induced subgraphs, as containing any of these would require that either $G$ or $\overline{G}$ not be chordal.

$\Longleftarrow$ Suppose $G$ does not contain a $2K_2$, $C_4$, or $C_5$ as an induced subgraph. Because any $C_n$, $n > 5$ contains a $2K_2$ as an induced subgraph, $G$ must not contain any induced cycles larger than $C_3$. Therefore, $G$ is chordal. Similarly, because $\overline{C_4} = 2K_2$ and $\overline{C_5} = C_5$, $\overline{G}$ also does not contain $2K_2$, $C_4$, or $C_5$ as induced subgraphs. By the same argument, $\overline{G}$ is chordal. Therefore, because $G$ and $\overline{G}$ are chordal, $G$ must be split.

# 3 Optimization Problems

## 3.1 Runtimes and $O(n)$

Split graphs afford significant computational speed-ups on traditionally difficult algorithmic problems in graph theory. Computational problems can be categorized by their runtime as a function of their input. This is represented by something called *big-O notation*. Take, as a simple example, the problem of finding an item in a list of length $n$. Supposing that this list is not ordered in any meaningful way, the best approach we have for finding our item is simply checking everything in the list until we find it. In the worst case scenario, this requires $n$ steps. We therefore say that this problem has a runtime of $O(n)$.

Runtimes are a useful metric for understanding if a problem is computationally tractable. While computers are capable of running millions of computations a second, algorithms whose runtimes grow exponentially as a function of their inputs can present a problem. Suppose, for example, that you had an algorithm whose runtime was $O(10^n)$. This would probably be fine for smaller $n$, but once you get to around $n = 10$ things will start to slow down. Even worse, your program will run ten times slower every time you increase the input size by 1. It is for this reason that computer scientists and mathematicians distinguish between programs that run in *polynomial time* ($n$ is some polynomial function) and *non-polynomial time* ($n$ is an exponential or a factorial or something worse). The problem of $P = NP$ posits that there exists a polynomial time algorithm for any computational problem.

Many problems in graph theory require non-polynomial time to solve on arbitrary graphs. These include finding a graph's maximum clique, maximum independent set, minimum clique covering, and problems of coloring. In the following sections, we will present solutions to these problems on split graphs that demonstrate a significant improvement in algorithmic runtime.

## 3.2 Maximum clique and maximum independent set

The following result from Hammer and Simeone [2] provides a characterization of a split graph $G$ in terms of its maximum clique $\omega(G)$ and minimum independent set $\alpha(G)$:

**Theorem 3.1** (Hammer and Simeone)**.** *Let $G$ be a split graph with clique $K$ and independent set $I$. Exactly one of the following statements is true:*

1. *$n(I) = \alpha(G)$ and $n(K) = \omega(G)$. In this case, $G = (K, I)$ is a unique partition of $G$ into a clique and independent set.*

2. *$n(I) = \alpha(G)$ and $n(K) = \omega(G) - 1$. In this case, there is an $x \in I$ such that $K \cup \{x\}$ is complete.*

3. *$n(I) = \alpha(G) - 1$ and $n(K) = \omega(G)$. In this case, there is a $y \in K$ such that $I \cup \{y\}$ is independent.*

*Proof:* The following proof is due to Golumbic [**?**]. Since a clique and an independent set can have at most one common vertex, $\omega(G) + \alpha(G)$ must be either $n(G)$ or $n(G) + 1$. We proceed in cases:

1. Suppose $\omega(G) + \alpha(G) = n(G)$. This corresponds to statement 1. We show that there is no other partition of $G$ into $K$ and $I$. Suppose otherwise. Let $G = (K', I')$ where $K'$ is a clique and $I'$ is an independent set. Let $\{x\} = I \cap K'$ and let $\{y\} = I' \cap K$. We consider two cases:

   (a) Suppose $xy \in E(G)$. Then $K \cup \{x\}$ is a clique of size $\omega(G) + 1$. This is not possible.

   (b) Suppose $xy \notin E(G)$. Then $I \cup \{y\}$ is an independent set of size $\alpha(G) + 1$. This is not possible.

   Therefore, $I \cap K' = I' \cap K = \varnothing$ and $G = (K, I)$ is unique.

2. Suppose $\omega(G) + \alpha(G) = n(G) + 1$. This corresponds to statements 2 and 3. For statement 2, let $n(K) = \omega(G) - 1$ and $n(I) = \alpha(G)$. Let $K'$ be a clique of size $\omega(G)$. Note that $K' \cup I$ is non-empty, since $K' > K$ and $G = (K, I)$. Therefore, $K' \cup I = \{x\}$ and $K' = K \cup x$. The same argument follows symmetrically for statement 3.

From theorem 2.1, we observe that it takes linear time to identify whether or not a graph is split. Note also that theorem 2.1 provides a candidate clique and independent set. From theorem 3.1, it clearly takes linear time to identify whether this clique is maximum (since we must simply iterate over the independent set). Therefore, the problem of finding a maximum clique is linear. Similarly, it takes linear time to identify whether the candidate independent set is maximum. Therefore, finding a maximum independent set is also solvable in linear time on split graphs.

### 3.3 Minimum clique covering

In [4], Gavril presents a construction for the clique cover of a chordal graph $G$. To understand this result, we must first present some definitions:

**Definition 3.1.** *An orientation of a graph's edges is called an R-orientation if:*

1. *The resulting digraph has no directed cycles.*

2. *If $b \to a$ and $c \to a$, then either $b \to c$ or $c \to b$.*

The following theorem is due to Rose [5]:

**Theorem 3.2.** *A finite graph $G$ is chordal if and only if it has an R-orientation.*

In the proof of this theorem, Rose develops an algorithm for constructing an $R$-orientation of a chordal graph. We shall not present the proof here, but we will note that the $R$-orientation of a chordal graph is such that each vertex is labeled $1, ..., n$ and each edge between two vertices is directed from low to high. We now present Gavril's construction for the minimum clique cover of a chordal graph $G$ [4].

Let $G$ be an $R$-oriented chordal graph. Let $i$ be a vertex in $G$. Let $J_i$ be the set of all vertices $j$ such that $j \to i$. We inductively define the following sequence of vertices $n_1, n_2, ..., n_t$. Let $n_1 = n$. Let $n_k$ be the largest vertex smaller than $n_{k-1}$ not in $J_{n_1} \cup ... \cup J_{n_{k-1}}$. Any vertex smaller than $n_t$ is in $J_{n_1} \cup ... \cup J_{n_t}$. From this definition, we have that $\{n_1, ..., n_t\} \cup J_{n_1} \cup ... \cup J_{n_t} = V(G)$.

We observe that $\{n_1, ..., n_t\}$ is an independent set of size $t$, and that therefore any minimum clique cover must contain at least $t$ cliques. We also observe that every $S_{n_i} = J_{n_i} \cup \{n_i\}$ is a clique and the set

$(S_{n_1}, ..., S_{n_t})$ covers $G$. We therefore conclude that the minimum clique cover is of size $t$ and is given by $(S_{n_1}, ..., S_{n_t})$.

To calculate the runtime of an algorithm designed to find a minimum clique cover of a chordal graph, note that it takes $(n-k-1)(k-1)$ steps to find the largest vertex smaller than $n_{k-1}$ not in $J_{n_1} \cup ... \cup J_{n_{k-1}}$. Since this operation must be done for $k$ vertices, we have:

$$\sum_{k=1}^{t}(n-k-1)(k-1) = \frac{t(t-1)(3n-2t+1)}{6}.$$

Since split graphs are a subset of chordal graphs, we conclude that a minimum clique covering on a split graph can be found in at least polynomial time.

### 3.4 Graph coloring

In split graphs, the problem of vertex coloring is reduced to the problem of finding the maximum clique. Let $G$ be a split graph with maximum clique $K$ and independent set $I$. We observe that all vertices in $I$ may be colored using the same colors as vertices in $K$. Suppose otherwise. Then there must be a vertex in $I$ adjacent to all vertices in $K$. However, this implies that $K$ is not maximum. Therefore, coloring $I$ does not require additional colors. Since a complete graph on $k$ vertices is $k$-colorable, we conclude that a split graph is $\omega(G)$-colorable ($\chi(G) = \omega(G)$).

As we observed above, the algorithm for finding a maximum clique in $G$ is linear on the size of $G$. Once a maximum clique is identified, the algorithm can simply greedy color first $K$, then $I$. Since $I$ is independent, the greedy coloring will use no colors other than those used to color $K$. Since the greedy coloring algorithm runs in linear time, we conclude that the minimum vertex coloring on $G$ may be computed in linear time.

## 4 References


[1] Foldes, S. and Hammer, P.L. (1977). *Split graphs.* Proceedings of the 8th South-Eastern Conference on Combinatorics, Graph Theory and Computing. 311–315.

[2] Hammer, P.L., Simeone, B (1981). *The splittance of a graph.* Combinatorica **1**: 275-284. https://doi.org/10.1007/BF02579333

[3] Wood, David. (2004). *Characterisations of Intersection Graphs by Vertex Orderings.* Computing Research Repository - CORR. 34.

[4] Gavril, F. (1972). *Algorithms for Minimum Coloring, Maximum Clique, Minimum Covering by Cliques, and Maximum Independent Set of a Chordal Graph.* SIAM Journal on Computing, 1(2), pp. 180-187. doi:10.1137/0201013

[5] Rose, D. J. (1970). *Triangulated graphs and the elimination process.* J. Math. Anal. Appl., 32, pp. 597-609.

[6] Merris, Russell (2003). *Split graphs.* European Journal of Combinatorics. **24**: 413-430. doi:10.1016/S0195-6698(03)00030-1