Rings which are generated by their units: 
a graph theoretical approach

H.R. Maimani, M.R. Pournaki and S. Yassemi

1 Introduction

The study of rings which are generated additively by their units seems to have arisen in 1953–1954 when Wolfson [13] and Zelinsky [14] proved, independently, that if $V$ is a finite or infinite dimensional vector space over a division ring $D$, then every linear transformation is the sum of two nonsingular linear transformations unless $\dim V = 1$ and

Der folgende Beitrag behandelt eine Strukturfrage zur Theorie endlicher kommutativer Ringe. Solche Ringe sind beispielsweise durch die Restklassenringe $\mathbb{Z}/n\mathbb{Z}$ ($n \in \mathbb{N}$) oder durch direkte Produkte solcher gegeben. Ein Element $u$ eines kommutativen Ringes $R$, das in $R$ ein multiplikatives Inverses $u^{-1}$ besitzt, wird Einheit genannt. Man sagt, dass der Ring $R$ durch Einheiten erzeugt ist, wenn sich jedes Element von $R$ als Summe von Einheiten darstellen lässt. In diesem Beitrag wird unter Verwendung graphentheoretischer Methoden in elementarer Weise gezeigt, dass ein endlicher kommutativer Ring $R$ mit von 0 verschiedenen Einselement genau dann durch Einheiten erzeugt ist, wenn es in $R$ kein Ideal $I$ mit Faktorring $R/I \cong \mathbb{Z}/2\mathbb{Z} \times \mathbb{Z}/2\mathbb{Z}$ gibt.
This implies that the ring of linear transformations $\text{End}_D(V)$ is generated additively by its units. In fact, every element of $\text{End}_D(V)$ is the sum of two units except for one obvious case when $V$ is a one dimensional vector space over $\mathbb{Z}_2$. Wolfson’s and Zelinsky’s result caused quite a bit of interest in the study of rings that are generated by their units.

In 1958, Skornyakov [8, p. 167, Problem 31], posed the problem of determining which regular rings are generated by their units. More precisely, he asked: Is every element of a von Neumann regular ring, which cannot have $\mathbb{Z}_2$ as a quotient, a sum of units? This question of Skornyakov was answered negatively by Bergman in 1977 (see [5] which is a significant contribution to the theory of von Neumann regular rings). Bergman constructed a von Neumann regular algebra in which not all elements are sums of units.

In 1968, while apparently unaware of Skornyakov’s book, Ehrlich [2] produced a large class of regular rings generated by their units. He proved that if $R$ is a ring such that 2 is a unit and for every $a \in R$ there exists a unit $u \in R$ such that $aua = a$, then every element of $R$ is the sum of two units.

In 1974, Raphael [7] launched a systematic study of rings generated by their units, which he calls $S$-rings.

Finally, in 1976, Fisher and Snider [3] proved that if $R$ is a von Neumann regular ring with primitive factor rings artinian and 2 is a unit, then every element of $R$ can be expressed as the sum of two units.

In 1998, Wolfson’s and Zelinsky’s result was reproved by Goldsmith, Pabst and Scott where they remarked that this result can hardly be new but they were unable to find any reference to it in the literature (see [4]). Interest in this topic increased recently after they defined the unit sum number in [4].

For additional historical background the reader is referred to the paper [10], which also contains references to recent work in this area. Also see [9] for a survey of rings which are generated by their units.

The purpose of this note is to give an elementary proof of Theorem 1.1. The proof uses graph theory, and offers, as a byproduct, that if $R$ is a finite commutative ring with nonzero identity which is generated by its units, then every element of $R$ can be written as a sum of at most three units.

**Theorem 1.1 ([7, Corollary 7]).** Let $R$ be a finite commutative ring with nonzero identity. Then $R$ is generated by its units if and only if $R$ cannot have $\mathbb{Z}_2 \times \mathbb{Z}_2$ as a quotient.

### 2 Basic notation and properties of graphs

In this section we introduce some notation and definitions of graphs that will be used throughout the note. We also state and prove Lemmas 2.1 and 2.2 which are required in Section 3. Here, by a graph $G$ we mean a finite undirected graph without loops and multiple edges (unless otherwise specified). The reader is referred to [1] and [12] for a fuller treatment of the subject.

For a graph $G$, let $V(G)$ denote the set of vertices. Let $G$ be a graph and suppose $x, y \in V(G)$. We recall that a walk between $x$ and $y$ is a sequence $x = v_0, e_1, v_1, \ldots, e_k, v_k = y$. 

of vertices and edges of $G$, denoted by

$$x = v_0 \xrightarrow{e_1} v_1 \rightarrow \ldots \rightarrow v_k = y,$$

such that for every $i$ with $1 \leq i \leq k$, the edge $e_i$ has endpoints $v_{i-1}$ and $v_i$. Also a path between $x$ and $y$ is a walk between $x$ and $y$ without repeated vertices. The number of edges in a walk (counting repeats) or a path is called its length.

For the proof of Theorem 1.1 we need the following well-known fact. We state and prove it here for the convenience of the reader.

**Lemma 2.1.** Let $x$ and $y$ be distinct vertices of a graph $G$. If there is a walk between $x$ and $y$ then there is also a path between $x$ and $y$.

**Proof.** By assumption there is a walk between $x$ and $y$ and so we may select a walk $W : x = v_0 \xrightarrow{e_1} v_1 \rightarrow \ldots \rightarrow v_k = y$ of minimal length $k$ between $x$ and $y$. If $W$ is not a path, select a vertex that appears twice, say $v_i = v_j$ where $i < j$. Consider

$$W' : x = v_0 \xrightarrow{e_1} v_1 \rightarrow \ldots \rightarrow v_i \xrightarrow{e_{i+1}} v_{j+1} \rightarrow \ldots \rightarrow v_k = y.$$

Then $W'$ is a walk between $x$ and $y$ with length shorter than $k$, a contradiction. Therefore $W$ is a path between $x$ and $y$. \qed

A graph $G$ is called **connected** if for all vertices $x$ and $y$ there exists a path between $x$ and $y$. Otherwise, $G$ is called **disconnected**.

A **bipartite** graph is one whose vertex-set is partitioned into two (not necessarily nonempty) disjoint subsets, called **parts**, in such a way that the two end vertices for each edge lie in distinct parts. Among bipartite graphs, a **complete bipartite** graph is one in which each vertex is joined to every vertex that is not in the same part.

Let $G_1$ and $G_2$ be two vertex-disjoint graphs. The **category product** of $G_1$ and $G_2$ is denoted by $G_1 \circ G_2$. That is, $V(G_1 \circ G_2) := V(G_1) \times V(G_2)$; two distinct vertices $(x, y)$ and $(x', y')$ are adjacent if and only if $x$ is adjacent to $x'$ in $G_1$ and $y$ is adjacent to $y'$ in $G_2$.

We now state and prove the following lemma which will be used in the proof of Theorem 1.1. A bipartite graph is **nontrivial** if both parts of its vertex set are nonempty. For more information on this lemma we refer the reader to [11].

**Lemma 2.2.** Let $G_1$ and $G_2$ be two bipartite graphs at least one of which is nontrivial. Then $G_1 \circ G_2$ is disconnected.

**Proof.** We assume that $G_2$ is nontrivial. Thus $G_1$ is partitioned into two disjoint subsets $X_1$ and $Y_1$ as well as $G_2$ into two disjoint subsets $X_2$ and $Y_2$ in such a way that $|X_1| \geq 1$, $|X_2| \geq 1$ and $|Y_2| \geq 1$. Choose $a \in X_1$, $b \in X_2$ and $c \in Y_2$. We claim that there is no path
between \((a, b)\) and \((a, c)\) in \(G_1 \times G_2\). In order to do this, suppose in contrary that, there is a path \(P\) between \((a, b)\) and \((a, c)\) in \(G_1 \times G_2\):

\[
P : (a, b) \rightarrow (a_1, b_1) \rightarrow (a_2, b_2) \rightarrow \ldots \rightarrow (a_{n-1}, b_{n-1}) \rightarrow (a, c).
\]

We now obtain the walk \(\hat{W}\) in \(G_1\) and the walk \(\tilde{W}\) in \(G_2\) both with length \(n\):

\[
\hat{W} : a \rightarrow a_1 \rightarrow a_2 \rightarrow \ldots \rightarrow a_{n-1} \rightarrow a,
\]

\[
\tilde{W} : b \rightarrow b_1 \rightarrow b_2 \rightarrow \ldots \rightarrow b_{n-1} \rightarrow c.
\]

The existence of the walk \(\hat{W}\) implies that \(n\) is even while the existence of the walk \(\tilde{W}\) implies that \(n\) is odd, a contradiction. Thus there is no path between \((a, b)\) and \((a, c)\) in \(G_1 \times G_2\), which implies that \(G_1 \times G_2\) is disconnected.

Let us consider yet a few more definitions required for a complete understanding of the next section. For a graph \(G\) and vertices \(x\) and \(y\) of \(G\), the distance \(d(x, y)\) is the number of edges in a shortest path between \(x\) and \(y\). If there is no path between \(x\) and \(y\) then we write \(d(x, y) = \infty\). We recall that the largest distance among all distances between pairs of the vertices of a graph \(G\) is called the diameter of \(G\) and is denoted by \(\text{diam}(G)\). Finally, for a given vertex \(x \in V(G)\), the neighbor set of \(x\) is the set \(N_G(x) := \{v \in V(G) \mid v\) is adjacent to \(x\}\). Moreover, if \(G\) has a loop at vertex \(x\), then we always assume that \(x \in N_G(x)\).

### 3 Proof of Theorem 1.1

In this section, using the results presented in Section 2, we are able to prove Theorem 1.1. Let \(R\) be an arbitrary finite associative ring \(R\) with nonzero identity, say 1, which is preserved by homomorphisms and inherited by subrings. Let \(U_R\) be the set of units of \(R\). We attach a graph to \(R\), denoted by \(G_R\), based on the elements and units of \(R\). This graph is obtained by letting \(R\) be the set of vertices and defining distinct vertices \(x\) and \(y\) to be adjacent if and only if \(x + y \in U_R\). If we omit the word “distinct” in the definition of \(G_R\), we obtain the graph \(\overline{G}_R\); this graph may have loops. Note that if \(2 \notin U_R\), then \(\overline{G}_R = G_R\).

The graphs in Fig. 1 are the graphs attached to the rings indicated.

It is easy to see that, for given rings \(R\) and \(S\), if \(R \cong S\) as rings, then \(G_R \cong G_S\) as graphs. Also we have \(G_R \times G_S \cong G_{R \times S}\).

In Fig. 2 we illustrate these points for the direct product of the rings \(\mathbb{Z}_2\) and \(\mathbb{Z}_3\).

We need the following result, which is useful in the sequel.

**Lemma 3.1.** Let \(R\) be a finite commutative local ring with maximal ideal \(m\). Then the following statements hold:

(a) If \(|R/m| = 2\), then \(G_R\) is a complete bipartite graph.

(b) If \(|R/m| > 2\), then for every \(x, y \in R\) we have \(N_{G_R}(x) \cap N_{G_R}(y) \neq \emptyset\).
Proof. Part (a): Let $X = m$ and $Y = R \setminus m$. We have $V(G_R) = X \cup Y$ and $X \cap Y = \emptyset$. Therefore $X$ and $Y$ partition $V(G_R)$ into two subsets. It is clear that no pair of distinct elements of $X$ are adjacent. We show that no distinct elements of $Y$ are adjacent. In order to do this, fix an element in $R \setminus m$, say $a$. By assumption we have $R = m \cup (m + a) = m \cup (m + (-a))$. Now for distinct elements $x$ and $y$ in $R \setminus m$, we may write $x = m + a$ and $y = m' - a$ where $m, m' \in m$. If $x + y \in U_R$, then we conclude that $m + m' \in U_R$. Therefore $m$ has a unit element and so $m = R$, a contradiction. Thus $x + y \notin U_R$, which implies that $x$ and $y$ are not adjacent. Therefore no distinct elements of $Y$ are adjacent. Hence $G_R$ is a bipartite graph.

Suppose that $x \in X$ and $y \in Y$ are given. If $x + y \notin U_R$, then $x + y \in X$ and so $y \in X$, a contradiction. Thus $x + y \in U_R$, which implies that $x$ and $y$ are adjacent. Therefore each vertex of $X$ is joined to every vertex of $Y$ and so $G_R$ is complete bipartite.

Part (b): By assumption we conclude that $|U_R| \geq 2|R|/3$. Suppose that $x$ is an arbitrary element of $R$ and fix it. There are two possibilities: either $2x \notin U_R$ or $2x \in U_R$. If $2x \notin U_R$, then $\overline{G_R}$ has no loop at vertex $x$. On the other hand, for every element $u - x$, where $u \in U_R$, we have $u - x \neq x$ and $u - x$ is adjacent to $x$ in $\overline{G_R}$. This implies that $\{u - x \mid u \in U_R\} \subseteq N_{\overline{G_R}}(x)$ and so $|N_{\overline{G_R}}(x)| \geq |U_R| \geq 2|R|/3$. If $2x \in U_R$, then $\overline{G_R}$ has a loop at vertex $x$. On the other hand, for every element $u - x$, where $u \in U_R \setminus \{2x\}$, we have $u - x \neq x$ and $u - x$ is adjacent to $x$ in $\overline{G_R}$. This implies that $\{u - x \mid u \in U_R \setminus \{2x\}\} \cup \{x\} = \{u - x \mid u \in U_R\} \subseteq N_{\overline{G_R}}(x)$ and so we have again
\[ |N_{G_{GR}}(x)| \geq |U_R| \geq 2|R|/3. \] Therefore, in both cases, we have \[ |N_{G_{GR}}(x)| \geq 2|R|/3. \]
Now, for every \( x, y \in R \),
\[
|N_{G_{GR}}(x) \cap N_{G_{GR}}(y)| = |N_{G_{GR}}(x)| + |N_{G_{GR}}(y)| - |N_{G_{GR}}(x) \cup N_{G_{GR}}(y)| \\
\geq (2|R|/3) + (2|R|/3) - |R| \\
= |R|/3 \\
> 0
\]
and so \( N_{G_{GR}}(x) \cap N_{G_{GR}}(y) \neq \emptyset \) as required. \( \square \)

Now let \( R \) be a finite commutative ring with nonzero identity and fix it. We want to prove \( R \) is generated by its units if and only if \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient. We start with the proposition below which contains a necessary and sufficient condition for \( G_{GR} \) to be connected.

**Proposition 3.2.** \( G_{GR} \) is connected if and only if \( R \) is generated by its units.

**Proof.** (\( \Longrightarrow \)) Suppose that \( a \in R \) is written by the sum of some units and \( b \in R \) is adjacent to \( a \) in \( G_{GR} \). Therefore \( a + b \in U_R \) and so we may write \( b = c - a \), for some \( c \in U_R \). Thus \( b \) is the sum of some units.

Now suppose that \( x \in R \) is given. Since \( G_{GR} \) is connected, there exists a path between \( x \) and 1 and, therefore, by the above observation we conclude that \( x \) is the sum of some units. This means that \( R \) is generated by its units.

(\( \Longleftarrow \)) Suppose that \( a \in R \). Since \( R \) is generated by its units, we may write \( a = u_1 + \ldots + u_k \), where \( u_i \in U_R \), \( 1 \leq i \leq k \). We now have the walk
\[
0 \xrightarrow{e_1} -u_1 \xrightarrow{e_2} u_1 + u_2 \xrightarrow{e_3} -u_1 - u_2 - u_3 \xrightarrow{e_4} u_1 + u_2 + u_3 + u_4 \\
\ldots \xrightarrow{e_k} u_1 + \ldots + u_k = a
\]
between 0 and \( a \), when \( k \) is even and the walk
\[
0 \xrightarrow{e_1} u_1 \xrightarrow{e_2} -u_1 - u_2 \xrightarrow{e_3} u_1 + u_2 + u_3 \xrightarrow{e_4} -u_1 - u_2 - u_3 - u_4 \\
\ldots \xrightarrow{e_k} u_1 + \ldots + u_k = a
\]
between 0 and \( a \), when \( k \) is odd.

This implies that for every \( x, y \in R \) there is a walk \( W_1 \) between \( x \) and 0 as well as a walk \( W_2 \) between 0 and \( y \). The walks \( W_1 \) and \( W_2 \) together form a walk \( W \) between \( x \) and \( y \). By using Lemma 2.1, we conclude that there is also a path \( P \) between \( x \) and \( y \), which implies the connectedness of \( G_{GR} \). \( \square \)

The following proposition contains another necessary and sufficient condition for \( G_{GR} \) to be connected.

**Proposition 3.3.** \( G_{GR} \) is connected if and only if \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient.
Proof. Every finite commutative ring with nonzero identity is isomorphic to a direct product of finite local rings (see [6, p. 95]). Therefore, we may write \( R \cong R_1 \times \ldots \times R_k \), where every \( R_i \) is a local ring with maximal ideal \( m_i \).

\((\implies)\) Suppose by contrary that \( R \) has \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient. This implies that for at least two \( i \), for example \( i = 1, 2 \), we have \( |R_i/m_i| = 2 \). Now part (a) of Lemma 3.1 implies that \( G_{R_1} \) and \( G_{R_2} \) are both bipartite. Thus by using Lemma 2.2, we conclude that \( G_{R_1} \times G_{R_2} \) is disconnected.

On the other hand, by the observation just before Lemma 3.1, we have

\[
G_R \cong \begin{cases} 
G_{R_1} \times G_{R_2} & \text{if } k = 2, \\
(G_{R_1} \times G_{R_2}) \times G_{R_1 \times \ldots \times R_k} & \text{if } k \geq 3.
\end{cases}
\]

But for \( i = 1, 2 \) we have \( 2 \notin U_{R_i} \) and so \( G_{R_i} \cong G_{R_i} \). Therefore we obtain

\[
G_R \cong \begin{cases} 
G_{R_1} \times G_{R_2} & \text{if } k = 2, \\
(G_{R_1} \times G_{R_2}) \times G_{R_1 \times \ldots \times R_k} & \text{if } k \geq 3.
\end{cases}
\]

Now the disconnectedness of \( G_{R_1} \times G_{R_2} \) implies that \( G_R \) is also disconnected. This contradiction shows that \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient.

\((\implies)\) By assumption \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient. This implies that for at most one \( i \), we have \( |R_i/m_i| = 2 \). There are the following cases to be considered:

1. \( |R_i/m_i| > 2 \) holds for every \( i \).
2. \( |R_i/m_i| > 2 \) holds for every \( i \) except one of them.

First, suppose that \( k = 1 \). In this case \( R \cong R_1 \) is a finite local ring with maximal ideal \( m_1 \) in such a way \( |R_1/m_1| = 2 \). Thus, if \( R \) is a field, then we have \( R \cong \mathbb{Z}_2 \) and so \( \text{diam}(G_R) = 1 \). If \( R \) is not a field, then by using part (a) of Lemma 3.1 we conclude that \( G_R \) is complete bipartite with \( |R| \geq 4 \) and so \( \text{diam}(G_R) = 2 \).

Second, suppose that \( k \geq 2 \). In this case we may assume that \( |R_1/m_1| = 2 \) and \( |R_i/m_i| > 2 \) for every \( i \) with \( 2 \leq i \leq k \). Suppose that \( x = (x_1, x_2, \ldots, x_k) \) and \( y = (y_1, y_2, \ldots, y_k) \) are arbitrary distinct elements of \( R_1 \times R_2 \times \ldots \times R_k \). If either \( x_1, y_1 \in m_1 \) or \( x_1, y_1 \notin m_1 \), then by the same argument as above, we obtain a path between \( x \) and \( y \) with length at
most 2. This implies that \( d(x, y) \leq 2 \). Now, we may assume that \( x_1 \in m_1 \) and \( y_1 \notin m_1 \).

For every \( i \) with \( 2 \leq i \leq k \), consider \( w_i \) as follows:

\[
 w_i = \begin{cases} 
 1 & \text{if } x_i \in m_i, \\
 0 & \text{if } x_i \notin m_i. 
\end{cases}
\]

On the other hand, since for every \( i \) with \( 2 \leq i \leq k \) we have \(|R_i/m_i| > 2\), by using part (b) of Lemma 3.1 we conclude that \( N_{GR_i}(w_i) \cap N_{GR_i}(y_i) \neq \emptyset \). Therefore we may choose \( z_i \in N_{GR_i}(w_i) \cap N_{GR_i}(y_i) \). Thus we have the following walk in \( GR_1 \times R_2 \times \ldots \times R_k \):

\[
 (x_1, x_2, \ldots, x_k) \xrightarrow{c_1} (y_1, w_2, \ldots, w_k) \xrightarrow{c_2} (x_1, z_2, \ldots, z_k) \xrightarrow{c_3} (y_1, y_2, \ldots, y_k).
\]

This implies that \( d(x, y) \leq 3 \). Therefore, for every distinct \( x, y \in R_1 \times R_2 \times \ldots \times R_k \) we have \( d(x, y) \leq 3 \) and so \( \text{diam}(GR) = \text{diam}(GR_1 \times R_2 \times \ldots \times R_k) \leq 3 \).

Therefore in both cases we have \( \text{diam}(GR) \leq 3 \). Thus every two vertices of \( GR \) are joined by a path with length at most 3, which implies that \( GR \) is connected. \( \square \)

Propositions 3.2 and 3.3 imply that \( R \) is generated by its units if and only if \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient which completes the proof of Theorem 1.1. Our proof shows that if \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient, then not only \( GR \) is connected, but also \( \text{diam}(GR) \leq 3 \).

Therefore, we may state

**Corollary 3.4.** Let \( R \) be a finite commutative ring with nonzero identity. If \( R \) is generated by its units, or equivalently, \( R \) cannot have \( \mathbb{Z}_2 \times \mathbb{Z}_2 \) as a quotient, then every element of \( R \) can be written as a sum of at most three units.

**Acknowledgements.** The authors would like to thank the referee for his/her useful suggestions which led to an improvement of the present note. Also thanks are due to S.A. Seyed Fakhari from Sharif University of Technology for his useful comments.

**References**


H.R. Maimani  
Mathematics Section, Department of Basic Sciences  
Shahid Rajaee Teacher Training University  
Tehran, Iran  
*and*  
School of Mathematics  
Institute for Research and Fundamental Sciences (IPM)  
P.O.Box 19395-5746  
Tehran, Iran  
e-mail: maimani@ipm.ir

M.R. Pournaki  
Department of Mathematical Sciences  
Sharif University of Technology  
P.O. Box 11155-9415  
Tehran, Iran  
*and*  
School of Mathematics  
Institute for Research and Fundamental Sciences (IPM)  
P.O.Box 19395-5746  
Tehran, Iran  
e-mail: pournaki@ipm.ir

S. Yassemi  
School of Mathematics, Statistics and Computer Science  
College of Science, University of Tehran  
Tehran, Iran  
e-mail: yassemi@ipm.ir

The research of H.R. Maimani and M.R. Pournaki was in part supported by a grant from IPM (No. 87050213 and No. 87200111).  
The research of S. Yassemi was in part supported by a grant from the University of Tehran (No. 6103023/1/07).