# Equivalent and Strong Equivalent Graphs with Application in Clustering

**Silviu Bârzã**
Faculty of Mathematics and Informatics
*Spiru Haret* University
Bucharest, Romania
silviu.barza@gmail.com

**Abstract**
*Firstly I define equivalent and strong equivalent graphs in terms of isomorphic graphs and with additional use of walks matrix. Those definitions are used then to define a possible clustering process. In this process I do not use individual pieces of information, but set of information for which is defined a possible connection inside a given set of information.*
**Keywords:** *isomorphic graphs, walks matrix, clustering.*
**ACM/AMS Classification:** 05C30, 05C60, 05C75, 91G20

## 1. Introduction

Firstly I wish to present some of graph theory definitions. I make this because there are authors that used different definitions, starting with definition for graph. In such a definition, a graph $G$ is defined as $(V, E, f)$ where $V$ represent the vertexes, $E$ is the set of edges and $f$ is adjacency function.

The definition used by others (including myself) is the following:

**Definition 1.1** *A graph $G$ is a pair $(V, E)$ where $V$ is a finite set specifying the vertexes (generally considered as $V = \{1, 2, \ldots, n\}$) and $E$ is the set of unordered pairs of numbers from $V$ (generally presented as subsets of two values from $V$) named edges.*

In the above definition we do not assume that numbers from $V$ are different so we can not have $\{x, x\} \in E$. More then that, we do not assume that if $a$ and $b$ are two different pairs from $E$, then we can not have $a$ and $b$ formed with the same values $x$ and $y$ from $V$.

**Definition 1.2** *Let $G = (V, E)$ be a graph and $x$ and $y$ two different vertexes from $G$. We have a walk from $x$ to $y$ if and only if we have a sequence $x = v_1, v_2, \ldots, v_k = y$ of vertexes so that, for any $i = 1, 2, \ldots, k-1$, $\{v_i, v_{i+1}\}$ is an edge in $E$. If $x = y$, the walk is named loop. We specify the walk from $x$ to $y$ by $[x = v_1, v_2, \ldots, v_k = y]$.*

There are some matrices that can be associated with graphs. One of this matrix is adjacencies matrix, $A_G = (a_{ij})_{i,j=1,2,...,n}$, where $a_{ij} = 1$ if and only if there exist a walk from $i$ to $j$, otherwise $a_{ij} = 0$, and $n$ represents the number of vertexes. For this paper the next definition is important:

**Definition 1.3** *Let $G = (V, E)$ be a graph having $n$ vertexes. We define walks matrix as a $0, 1$ matrix $D_G = (d_{ij})_{i,j=1,2,...,n}$, where $d_{ij} = 1$ if and only if there exist a walk from $i$ to $j$ , otherwise $d_{ij} = 0$.*

I wish to remember here an algebraic result related to bijective functions of finite sets.

**Theorem 1.1** *Let $A$ and $B$ be two finite sets and $f : A \to B$ a function. Then the next affirmations are equivalent:*

1. *$f$ is bijective*
2. *$f$ is injective*
3. *$f$ is surjective.*

Another important result is:

**Propozition 1.1** *Let $A$ and $B$ be two finite sets. There exists a bijective function $f : A \to B$ if and only if $|A| = |B|$ ($A$ and $B$ have the same number of element, where $|X|$ represents the number of elements of $X$).*

## 2. **Isomorphic graphs**

In this section I present some results related to isomorphic graphs which is necessary to my presentation. Firstly we must know what it means isomorphic graphs.

**Definition 2.1** *Let $G = (V, E)$ and $H = (W, F)$ be two graphs. $G$ and $H$ are named isomorphic if and only if there exists a function $f : V \to W$ so that $f$ is bijective and $a = \{x, y\} \in E$ if and only if $f(a) = \{f(x), f(y)\} \in F$.*

One can demonstrate that:

**Note 2.1** *If $G = (V, E)$ and $H = (V, F)$ are two isomorphic graphs and $\sigma$ is a permutation of $V$ so that $F = \{\{\sigma(x), \sigma(y)\} \mid \{x, y\} \in E\}$, we write $H = \sigma(G)$.*

In clustering process will be useful to consider a result which is easily proofed using theorem 1.1 and proposition 1.1. So:

**Lemma 2.1** *Let $G = (V, E)$ and $H = (W, F)$ be two isomorphic graphs. Then $|V| = |W|$ and $|E| = |F|$.*

**Proofs** $G$ and $H$ are two isomorphic graphs if there exists a bijective function $f : V \to W$ so that $\{x, y\} \in E$ if and only if $\{f(x), f(y)\} \in F$.

Because $f : V \to W$ is a bijective function and $V$ and $W$ are two finite sets, from proposition 1.1 it follows that $|V| = |W|$.

Let $g : E \to F$ be a function define by $g(\{x, y\}) = \{f(x), f(y)\}$. If $\{x, y\}, \{s, t\} \in E$, $\{x, y\} \neq \{s, t\}$ if and only if $x \neq s$ or $y \neq t$ and so there are three cases:

1. $x \neq s$ and $y = t$. Because $f$ is an injective function and $x \neq s$, it follows that $f(x) \neq f(s)$, and from $y = t$ it follows that $f(y) = f(t)$. So

$$\{f(x), f(y)\} \neq \{f(s), f(y)\} = \{f(s), f(t)\},$$

   which means $g(x, y) \neq g(s, t)$.

2. $x = s$ and $y \neq t$. Similar with case 1 we have $g(x, y) \neq g(s, t)$.

3. $x \neq s$ and $y \neq t$. It is clear in this case that $g(x, y) \neq g(s, t)$.

In conclusion, it follows that $g$ is an injective function. Because $g$ is define on two finite sets, from theorem 1.1 results that $g$ is a bijective function, and using proposition 1.1 it follows that $|E| = |F|$.

## 3. **Equivalent and strong equivalent graphs**

If $G = (V, E)$ is a graph, there exists the function $1_V : V \to V$ (identity function on $V$, $1_V$ bijective, $1_V(x) = x$ for any $x \in V$ and It could be written that $\{1_v(x), 1_V(y)\} = \{x, y\}$ for any $\{x, y\} \in E$ and so $G$ is isomorphic with $G$ (we write $G \cong G$).

Let $G = (V, E)$ and $H = (W, F)$ be two graphs so that $G \cong H$, there exists a bijective function $f : V \to W$ so that $\{x, y\} \in E$ if and only if $\{f(x), f(y)\} \in F$. Because $f$ is bijective, it exists $f^{-1} : W \to V$, and $f^{-1}$ is also bijective. Considering $\{u, v\} \in F$, let $x = f^{-1}(u)$ and $y = f^{-1}(v)$ be two vertexes from $V$. Assuming that $\{x, y\} \notin E$ it follows that $\{f(x), f(y)\} \notin F$. But $f(x) = f(f^{-1}(u)) = u$ and $f(y) = f(f^{-1}(v)) = v$ and so it results that $\{u, v\} = \{f(x), f(y)\} \notin F$ which is a contradiction. In conclusion we have $\{x, y\} \in E$.

Let us consider now that $\{x, y\} \in E$. Because $x \in V$, there exists $u \in W$ so that $x = f^{-1}(u)$, and because $y \in V$, there exists $v \in W$ so that $y = f^{-1}(v)$. Because $\{x, y\} \in E$ it follows that $\{f(x), f(y)\} \in F$, but $f(x) = f(f^{-1}(u)) = u$ and $f(y) = f(f^{-1}(v)) = v$ and so $\{u, v\} \in F$. $\{x, y\}$ could be written as $\{f^{-1}(u), f^{-1}(v)\}$ and so, from $\{f^{-1}(u), f^{-1}(v)\} \in E$ it results that $\{u, v\} \in F$.

We have shown that $\{u, v\} \in F$ if and only if $\{f^{-1}(u), f^{-1}(v)\} \in E$ and so, because $f^{-1}$ is a bijective function it results that $H \cong G$ and in conclusion we have that $G \cong H$ if and only if $H \cong G$.

Let us consider three graphs $G = (V, E)$, $H = (W, F)$ and $K = (U, D)$ so that $G \cong H$ and $H \cong K$.

From $G \cong H$ it results that there exists $f : V \to W$, bijective so that $\{x, y\} \in E$ if and only if $\{f(x), f(y)\} \in F$.

From $H \cong K$ it results that there exists $g : W \to U$, bijective so that $\{u, v\} \in F$ if and only if $\{g(u), g(v)\} \in D$.

Let us define the function $h : V \to U$ as $h(x) = g(f(x)) = g \circ f(x)$. Because $f$ and $g$ are bijective functions it follows that $h = g \circ f$ is also a bijective function.

Because $\{x, y\} \in E$ if and only if $\{f(x), f(y)\} \in F$ and $\{u, v\} \in F$ if and only if $\{g(u), g(v)\} \in D$, if we consider $u = f(x)$ and $v = f(y)$ we can write that $\{f(x), f(y)\} \in F$ if and only if $\{g(f(x)), g(f(y))\} \in D$ and so it is true that $\{x, y\} \in E$ if and only if $\{g \circ f(x), g \circ f(y)\} \in D$.

In conclusion we have that $h$ is an isomorphic function from $G$ to $K$ and so we can write that $G \cong K$.

From the above proofs it follows that isomorphic property for graphs is the same as an equivalence relationship on graphs and we can give the next definition:

**Definition 3.1** *On the set of graphs, we say that two graphs $G$ and $H$ are equivalent if and only if, by definition, $G \cong H$ and we write $G \simeq_i H$.*

We already shown that if $G$ is a graph then $G \cong G$. In the same time if $L_G$ is the walks matrix of $G$ then we have $L_G = L_G$.

Above we also shown that if $G$ and $H$ are two graphs then $G \cong H$ if and only if $H \cong G$. If $L_G$ and $L_H$ are the walks matrices for $G$ and respectively $H$ and $L_G = L_H$ then we have also $L_H = L_G$.

Finally, let $G$, $H$ and $K$ be three graphs for which $G \cong H$ and $H \cong K$. Let $L_G$, $L_H$ and $L_K$ be the walks matrices respectively for $G$, $H$, and $K$ so that $L_G = L_H$ and $L_H = L_K$. We already shown that if $G \cong H$ and $H \cong K$ it follows that $G \cong K$. Also, from $L_G = L_H$ and $L_H = L_K$ it results that $L_G = L_K$.

The last three considerations allow us to give the next definition:

**Definition 3.2** *On the set of graphs we say that two graphs $G$ and $H$ are strong equivalent if and only if, by definition, $G \cong H$ and $L_G = L_H$, where $L_T$ is the walks matrix for graph $T$, and we write $G \approxeq_i H$.*

We call $\approxeq_i$ strong equivalence because we have:

**Lemma 3.1** *Let $G$ and $H$ be two graphs. If $G \approxeq_i H$ then $G \simeq_i H$.*

This affirmation is obvious because the difference between relationship $\simeq_I$ and relationship $\approxeq_i$ consists in adding a condition to make the transition from $\simeq_i$ to $\approxeq_i$.

**Note 3.1** If two graphs are isomorphic it is not necessary that those graphs have the same walks matrices.

To show that note 3.1 is true we consider the next example:

**Example 3.1** We consider the sets $V = W = \{1, 2, 3, 4, 5\}$ and

$$E = \{\{1, 2\}, \{3, 4\}, \{3, 5\}\}$$

and

$$F = \{\{1, 3\}, \{1, 5\}, \{2, 4\}\}.$$

So we define the graphs $G = (V, E)$ and $H = (W, F)$.

If we consider the permutation $\sigma$ which is a bijective function on $V = W$ given by

$$\sigma = \begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 4 & 2 & 1 & 3 & 5 \end{pmatrix}.$$

Because from $\{1, 2\} \in E$ we have $\{\sigma(1), \sigma(2)\} = \{4, 2\} = \{2, 4\} \in F$, from $\{3, 4\} \in E$ we have $\{\sigma(3), \sigma(4)\} = \{1, 3\} \in F$ and from $\{3, 5\} \in E$ we have $\{\sigma(3), \sigma(5)\} = \{1, 5\} \in F$ it results that $G \cong H$ and so we have $G \simeq_i H$.

For walks matrix of $G$ we have

$$L_G = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 1 \end{pmatrix}$$

and the walks marix of $H$ is

$$L_H = \begin{pmatrix} 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 1 \end{pmatrix}.$$

We can observe that $L_G \neq L_H$ and so it is not true that $G \approxeq_i H$.

## 4. Clustering using equivalent and strong equivalent graphs

I suppose that, by observations, it results sets of information. Inside each set, the information are connected on one characteristic and this allow us to associate a graph in which vertexes are the points of data from collected information and the edges are the known connection between pieces of information from the set.

To formalize, we can consider a collection of observations $E_1, E_2, \ldots, E_k$, where for any $i$, $1 \leq i \leq k$, $E_i$ is a set of information formed by individual date $d_{i,1}, d_{i,2}, \ldots, d_{i,m_i}$ and is known that there exists connection between $d_{i,x}$ and $d_{i,y}$ for some $x, y \in \{1, 2, \ldots, m_i\}$.

With this considerations, for any $i$, $1 \leq i \leq k$, we consider the points $p_{i,1}, p_{i,2}, \ldots, p_{i,m_i}$ as vertexes of a graph $G_i = (V_i, F_i)$ and so

$$V_i = \{p_{i,1}, p_{i,2}, \ldots, p_{i,m_i}\}.$$

To define $F_i$ we consider that if there exists a connection between $d_{i,x}$ and $d_{i,y}$, then we have an edge between $p_{i,x}$ and $p_{i,y}$, so $\{p_{i,x}, p_{i,y}\} \in F_i$.

From the above process we associate to the collection of sets of information a collection of graphs $G_1, G_2, \ldots, G_k$ and we wish to use the definitions for equivalent or strong equivalent graph to realize a clustering for the collection of information $E_1, E_2, \ldots, E_k$.

Before to propose a clustering process let us see what elements characterized the clusters.

Firstly, to cluster on base of isomorphic graphs, means that the graphs associated with sets of information are tested to isomorphic property and so to consider the results given in lemma 2.1. Even if the clustering is made with $\simeq_i$ relationship or with $\approxeq_i$ relationship, in the same cluster will enter sets of information with isomorphic graphs associated and so the graphs which have the same number of vertexes and the same number of edges.

The above considerations let us say that the cluster is characterized by $n$, $m$ and $G = (V, E)$, where $G$ is a graph associated with one of the set of information from cluster, $n$ is the number of vertexes in $G$ ($n = |V|$), and $m$ is the number of edges in $G$ ($m = |E|$) and, in the same time, the number of known connection between individual pieces of information in a set of information.

Additional, if we use $\approxeq_i$ relationship, then we have to add a $\{0, 1\}$-matrix which represents the walks matrix for the graph associated with sets of information from the cluster.

Summarizing, if clustering is made on the base of $\simeq_i$ relationship, a cluster $C_i$ is characterized by

$$char(C_i) = [n_i, m_i, G_i = (V_i, F_i)],$$

and if clustering is made using $\approxeq_i$ relationship, then a cluster $C_i$ is characterized by

$$char(C_i) = [n_i, m_i, G_i = (V_i, F_i), L_i = L_{G_i}].$$

To indicate a possible clustering process I consider that, at a given moment, I already have a sets of clusters $C_1, C_2, \ldots, C_p$, characterized by $char(C_i)$ for any $i$, $1 \leq i \leq p$, and it is necessary to put a new set on information, $E$ in a cluster.

The process I propose is:

1. It is generated the graph $G = (V, F)$ associated with $E$.

2. If does not exists a cluster $C_i$ with $n_i = |V|$ and $m_i = |F|$, for $1 \leq i \leq p$, then we go to step 5.

3. Let $C_{i_1}, C_{i_2}, \ldots, C_{i_r}$ be the clusters for which $n_{i_j} = |V|$ and $m_{i_j} = |F|$, $1 \leq j \leq r$, $i_1, i_2, \ldots, i_r \in \{1, 2, \ldots, p\}$.

4. We test if $G$ and $G_{i_j}$ are isomorphic graphs, $1 \leq j \leq r$. If such a test is true then if $G_k \cong G$ we go to step 6, otherwise we continue with step 5.

From this step on we have to consider which relationship is used in clustering process.

If $\simeq_i$ relationship is used, then we continue with:

5. It is created a new cluster $C_{p+1}$ with $n_{p+1} = |V|$, $m_{p+1} = |F|$ and $G_{p+1} = G(V, F)$; we replace $p$ with $p + 1$ and stop.

6. We place $E$ in cluster $G_k$ and stop.

If $\approxeq_i$ relationship is used, then we continue with:

5. It is created a new cluster $C_{p+1}$ with $n_{p+1} = |V|$, $m_{p+1} = |F|$ and $G_{p+1} = G(V, F)$; we compute $L_{p+1} = L_G$ as walks matrix for $G$; we replace $p$ with $p + 1$ and stop.

6. Let $C_{j_1}, C_{j_2}, \ldots, C_{j_s}$ be the clusters for which $n_{j_t} = |V|$ and $m_{j_t} = |F|$ and $G \cong G_{j_t}$, $1 \leq t \leq s$, $j_1, j_2, \ldots, j_s \in \{i_1, i_2, \ldots, i_r\}$.

7. We compute $L_G$, the walks matrix for $G$

8. If there exists $k$, $k \in \{j_1, j_2, \ldots, j_s\}$, so that $L_G = L_{G_k}$ then we continue with step 9, otherwise we go to step 5.

9. We place $E$ in cluster $G_k$ and stop.

Finally I will give some notes on the process specified above.

I place the generation of graph associated with new set of information in first step because is a useful piece of information in the rest of the process and so I avoid that the same operation to be realized in more then one steps.

I consider that is necessary to make a clear separation between the use of $\simeq_i$ relationship and the use of $\approxeq_i$ relationship. So I place general steps (which are used for both relationships) in first four steps and then the steps which are specific for a given relationship (steps 5 and 6 for $\simeq_i$ relationship, steps from 5 to 9 for $\approxeq_i$ relationship).

As regard the complexity of this process, this characterization will be given by the complexity of steps 4 because the operation to determine if two graphs are isomorphic is the most complex operation in the above process.

5. **Conclusion**

In this paper I wish only to propose a new approach for clustering, starting from isomorphic graphs and so I build a clustering process which can be very simple for those who work with graphs.

The goal of this paper is only to propose a clustering. This process need to be analyzed an I wish to do such a work in the future.

Finally I want to remark that all preliminary definitions and results are english version for those in ([2])

**References**


1. Bamdy, J.A., Murty U.S.R., *Graph Theory*, "Springer-Verlag", 2007.
2. Barza, S., and Morogan, L.M., *Algoritmica grafurilor*, "Editura Fundatiei Romania de Maine", Bucharest, 2008.
3. Berge, C., *Graph Theory and Application*, romanian edition, "Editura Tehnica", Bucharest, 1971.
4. Popescu, D.R., *Combinatorica si teoria grafurilor*, "Editura Societatii de Stiinte matematice din Romania", Bucharest, 2005.
5. Tomescu, I., *Combinatorica si teoria grafurilor*, "Editura Universitatii Bucuresti", Bucharest, 1990.