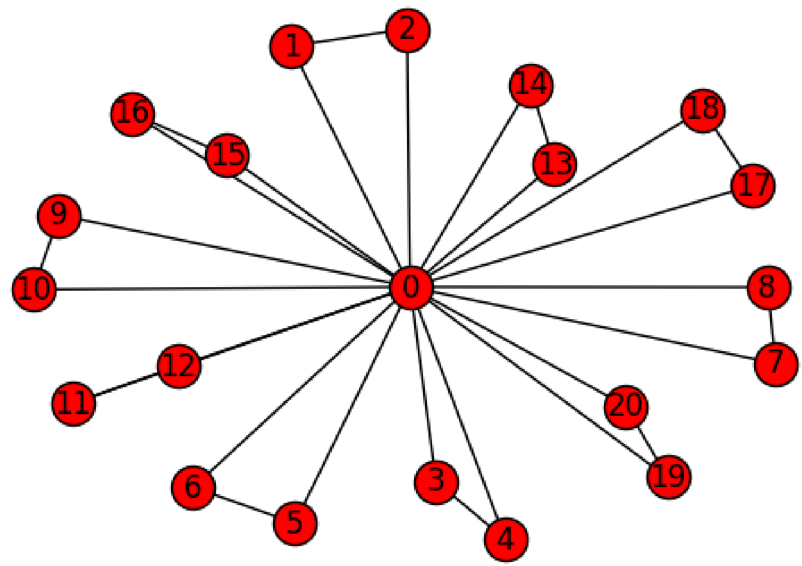
# Challenge Problem 5

## The clustering coefficient of the Dutch Windmill graph ¶

### Jacob Craffey

In [43]:
```
import warnings
import pandoc
warnings.filterwarnings('ignore')
import networkx as nx
import matplotlib.pyplot as plt

# Generate the Dutch windmill graph D(m)
# This uses a dictionary representation.

def windmill(m):
    # Start by letting 0 be the common node in the "center"
    # and connecting it to all the other nodes.
    # This uses a "list comprehension".
    dw = {0:[i for i in range(1,2*m+1)]}

    # Now connect 2 to 0 and 1, 4 to 0 and 3, 6 to 0 and 5, etc.
    for i in range(1,m+1):
        dw[2*i] = [0, 2*i-1]

    # Define the graph object and return it.
    return nx.Graph(dw)

# Prints windmill graph
def printGraph (n):
    windmill_graph = windmill(n)
    nx.draw(windmill_graph, with_labels=True)
    plt.show()
    print (nx.clustering(windmill_graph))
    print ()
    print ("Overall cluster coefficient with n=" + str(n) + ": " + str(nx.aver
age_clustering(windmill_graph)))

printGraph(10)
```

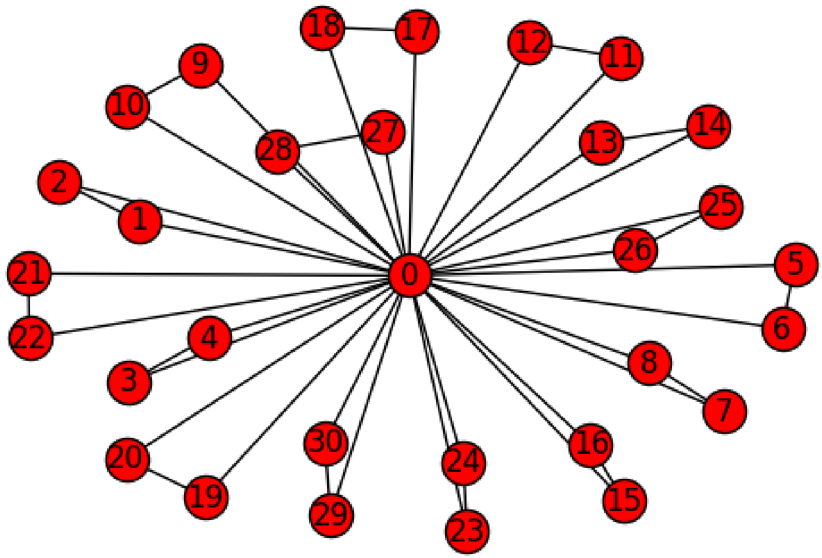



```
{0: 0.05263157894736842, 1: 1.0, 2: 1.0, 3: 1.0, 4: 1.0, 5: 1.0, 6: 1.0, 7:
 1.0, 8: 1.0, 9: 1.0, 10: 1.0, 11: 1.0, 12: 1.0, 13: 1.0, 14: 1.0, 15: 1.0, 1
6: 1.0, 17: 1.0, 18: 1.0, 19: 1.0, 20: 1.0}

Overall cluster coefficient with n=10: 0.9548872180451129
```

## Decomposition of the Dutch Windmill Graph

If one were to observe the average cluster coefficient of a Dutch Windmill Graph $G_n$.

In [44]:
```
printGraph(15)
```




```
{0: 0.034482758620689655, 1: 1.0, 2: 1.0, 3: 1.0, 4: 1.0, 5: 1.0, 6: 1.0, 7:
 1.0, 8: 1.0, 9: 1.0, 10: 1.0, 11: 1.0, 12: 1.0, 13: 1.0, 14: 1.0, 15: 1.0, 1
6: 1.0, 17: 1.0, 18: 1.0, 19: 1.0, 20: 1.0, 21: 1.0, 22: 1.0, 23: 1.0, 24: 1.
0, 25: 1.0, 26: 1.0, 27: 1.0, 28: 1.0, 29: 1.0, 30: 1.0}

Overall cluster coefficient with n=15: 0.9688542825361514
```

# Mathematical Conjecture

In this special graph, one key observation is that the Dutch Windmill Graph is composed of two distinct types of nodes. The center node that connects the $C_3$ graphs, and the non-center nodes that compose the $C_3$ graphs.

From the Dutch Windmill Graphs shown above, we are able to see the cluster coefficient for each node. One observation to take away from this information is that the cluster coefficient for the non-center nodes is always equal to 1.

The cluster coefficient for the center node for the graph $D_n$ can be expressed by:

$$\frac{1}{2n-1}$$

The cluster coefficient can be expressed in such a way because for each $n$ cycles the center connects. There are always 2 edges that are connected to the center node to complete the $C_3$. Furthermore, we subtract 1 from $2n$ because the center node should not count itself.

Then, we add the center node's cluster coefficient by the number of nodes there are in the Dutch Windmill Graph. Which then brings us to the equation:

$$\frac{1}{2n-1} + 2n$$

Since the non-center nodes each have a cluster coefficient of 1. The sum of the cycle's non-center nodes cluster coefficients can be represented as $2n$.

Finally, we divide all of the sum of coefficients by the number of nodes in the Dutch Windmill Graph. This formula has been given to us and is expressed as:

$D_m = 2m + 1$

**For all positive integers *n*, the overall clustering coefficient of $D_n$ is:**

$$\frac{\frac{1}{2n-1} + 2n}{2n+1}$$

# Proof

The purpose of the proof is to come up with a formula that is able to calculate the cluster coefficient of a Dutch Windmill Graph with *n* cycles. It is less complex and more efficient than finding the overall cluster coefficient using the original formula:

$$\langle C \rangle = \frac{1}{m} \sum_{i}^{m} \frac{2n_i}{k_i(k_i - 1)}$$

Where *n* represents the amount of neighboring connections to the node of node *i*, *k* represents the degree of the node *i*, and *m* represents the amount of nodes that compose the graph.

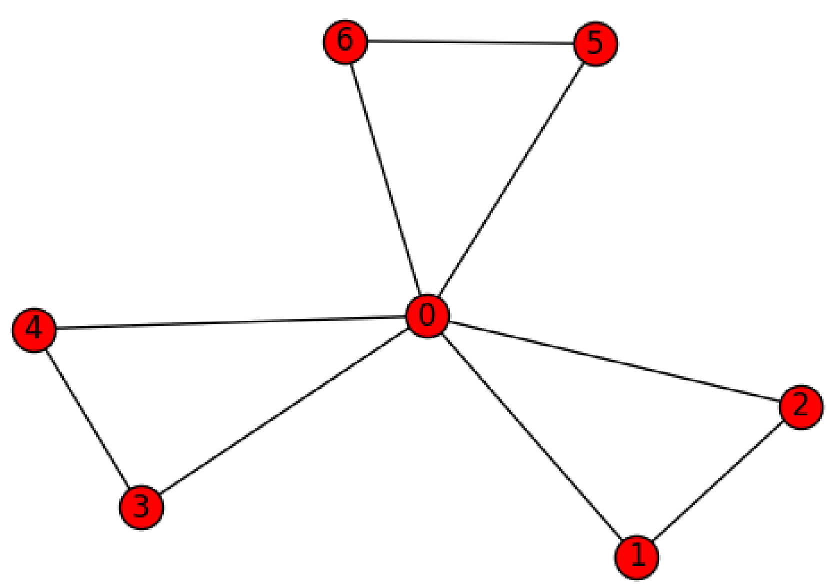
The structure of the conjecture consists of: the sum of the center node cluster coefficient and the non-center nodes' cluster coefficient. Divided by the amount of nodes that make up the graph.

To prove this conjecture, the formula is going to be broken apart into seperate pieces in order to explain why each part is significant and correct.

First, the center node's cluster coefficient in terms of *n*:

$$\frac{1}{2n - 1}$$

In [45]:
```
printGraph(3)
```




```
{0: 0.2, 1: 1.0, 2: 1.0, 3: 1.0, 4: 1.0, 5: 1.0, 6: 1.0}

Overall cluster coefficient with n=3: 0.8857142857142858
```

Suppose we have a Dutch Windmill Graph composed where *n* = 3. If we were to calculate the cluster coefficient of the center node with the original equation. The equation would be:

$$C_i = \frac{2(3)}{(6(6-1)} = \frac{1}{3}$$

From the graph shown above, some patterns to note.

1. There are always going to be *n* neighbor connections in a Dutch Windmill Graph with *n* cycles. These are represented as the outside edge that connects the two non-center nodes (eg: (1,2)).
2. The degree for the center node is always going to be 2*n*. This is because for every cycle *n* in the graph, there must be 2 edges to complete the cycle. With those observations, we are able to substitute them in the original equation.

$$\frac{1}{2n-1} = \frac{2n_i}{k_i(k_i-1)}$$

$$\frac{1}{2n-1} = \frac{2n}{2n(2n-1)}$$

$$\frac{1}{2n-1} = \frac{1}{2n-1}$$

**Cluster coefficient for center node is TRUE**.

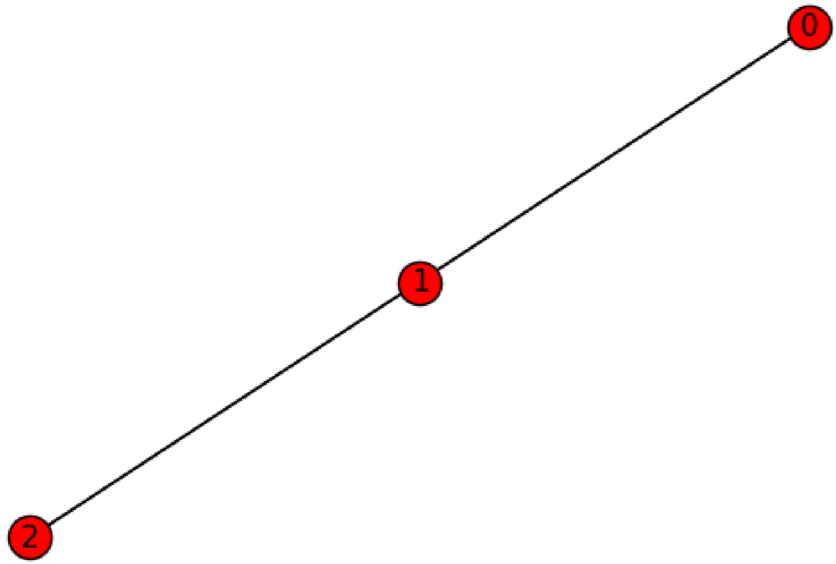
Next, we look at the cluster coefficient with the non-center nodes and prove that my conjecture is equal to the original cluster coefficient formula. A non-center node's cluster coefficient for a Dutch Windmill Graph in terms of *n* would be represented as:

$$n$$

Since there are 2 non-center nodes for every *n* cycle, the formula should now be represented as:

$$2n$$

```
In [46]: printGraph(1)
```



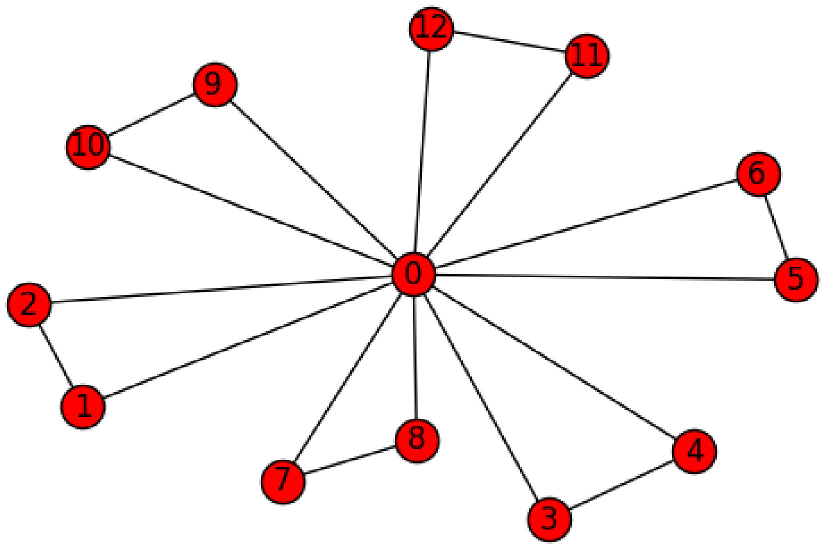

```
{0: 1.0, 1: 1.0, 2: 1.0}

Overall cluster coefficient with n=1: 1.0
```

Since Dutch Windmill Graph's are composed of *n* cycles of $C_3$. There is one clear pattern we can observe. Shown from the 1 $C_3$ shown above, we can see that every node's neighbor edge connections is n and every degree is 2. We can replace these values into the original.

$$n = \frac{2n_i}{k_i(k_i - 1)}$$
$$n = \frac{2(n)}{2(2-1)}$$
$$n = n$$

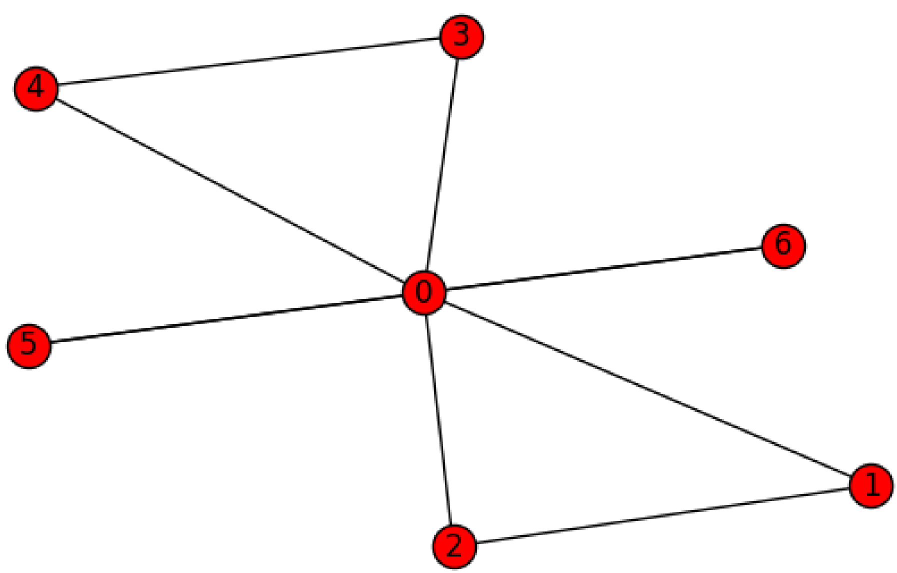
**Cluster coefficient for non-center nodes is TRUE**

In [47]:
```
printGraph(6)
printGraph(3)
```




{0: 0.09090909090909091, 1: 1.0, 2: 1.0, 3: 1.0, 4: 1.0, 5: 1.0, 6: 1.0, 7: 1.0, 8: 1.0, 9: 1.0, 10: 1.0, 11: 1.0, 12: 1.0}

Overall cluster coefficient with n=6: 0.93006993006993




{0: 0.2, 1: 1.0, 2: 1.0, 3: 1.0, 4: 1.0, 5: 1.0, 6: 1.0}

Overall cluster coefficient with n=3: 0.8857142857142858

Lastly, in order to get the average cluster coefficient of the graph, we must divide it by the number of nodes the graph is composed of.

From the graphs shown above where *n* = 6 and *n* = 3. We can see that each of them are composed of 13 and 7 nodes respectively. For every cycle *n* there are two nodes with the additional center node. This denotes an equation that can be expressed as:

$$2n+1$$

This replaces the (1/m) part of the equation of the original cluster coefficient.

The mathematical conjecture that has been found is proven to equal the overall cluster coefficient for Dutch Windmill Graph.

$$\frac{\frac{1}{2n-1}+2n}{2n+1} = \frac{1}{m}\sum_{i}^{m}\frac{2n_i}{k_i(k_i-1)}$$

■