Application of Graph Theory in Six Degrees of Separation with Small-World Network

Muhammad Daru Darmakusuma 13518057
Program Studi Teknik Informatika
Sekolah Teknik Elektro dan Informatika
Institut Teknologi Bandung, Jl. Ganesha 10 Bandung 40132, Indonesia
13518057@std.stei.itb.ac.id

***Abstract*—Seven and a half billion people live on this earth today. Most people have friends, 40 of them by average, and their friends have friends too with the same average number of it. With that information in mind, we can think that we can connect to a person we wanted to by connecting to our friends connections and their friends connections and so on. A small-world network is one of a few representation for the illustration before. A small-world network is one of a type for mathematical graph that most nodes aren't neighbors with one another, but the neighbors of given nodes likely can reach another neighbors with small hops or steps. The term of "small-world network" is also frequently refered as Watts-Strogatz toy network. With the knowledge of small-world network we can take a lot of application from it. Small-world network can be used in sociology, earth sciences, computing, even neural network in the brain. This paper will discuss the application of graph theory with small-world network in social connection or known as six degrees of separation.**

***Keywords*—Graph theory, Six degree of separation, Small-world network, Social connection.**

I. Introduction

In the late 1920s, a Hungarian writer wrote a short story, that titled *Láncszemek* or can be translated as *Chains or Chain-Links* in English. In the story, there are two character that believed that any individual on Earth can be connected with each other through a chain that not greater than five people. Then, in the 1950s, Ithiel de Sola Pool and Manfred Kochen wrote a manuscript called "Contacts and Influence", which expressed important ideas in social network and included a discussion of quantifiying the distance between people. To test the existence of short path connections between people, psychologist Stanley Milgram conducted a landmark studies at the 1960s on the small-world phenomenon in human social networks. Milgram sought to quantify the typical distance between actors in a social network and to show that one should expect it to be small.

In the experiment, Milgram sent 96 packages to people who lived in Omaha, NE, and USA that he selected randomly from a telephone directory. The package contains an official booklet that included the crest of Harvard University, the collage he was in. The information that included from Milgram's friends are their name, gender, and address. Each recipient was instructed to send the package to the person that they knew by the first-name basis who they felt would be socially closer to the target individual.

Total of the target received 18 out of the 96 packages. This rate of success was higher than expected and now, there is the modern version of this experiment by using e-mail with a smaller rate of success. Milgram asked the participants to record in the package each step of path and the average number of hops completed path is 5.9. This result led to popularization of the idea that there are no more then 6 steps between each pair of people in the world, which is encapsulated by the phrase "6 degrees of separation".

II. Basic Theory

A. Graph

Graphs are discrete structures consisting of vertices and edges that connect these vertices. There are different kinds of graphs, depending on whether edges have directions, whether multiple edges can connect the same pair of vertices, and whether loops are allowed[1]

Graphs formally denoted as

$$G = (V , E)$$

which consist *V*, a nonempty set of *vertices*, *nodes*, or *points* and *E*, a set of *edges*, *lines*, or *arcs*. Each edge has either one or two vertices that related with one another that is called *endpoints*. An edge can be said connected to its endpoints.

The vertices and the edges at a graph can be represented as different kind of variations, it is often shown as a relation of objects. The meaning of each graph can be differ depending the context of the graph itself.

In graph, the vertices often labelled with an alphanumerical form while the edges labelled with *e* that followed by the subscript of the labels which vertices it connects. Edges frequently denoted as

$$e = (u, v)$$

which consist *u* and *v*, the vertices that connected by *e* edge.

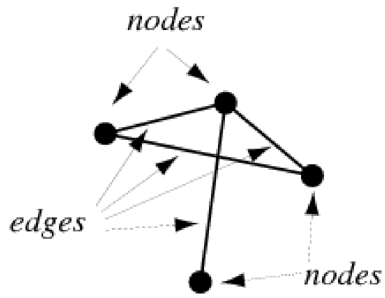

Fig 1. An example of a graph
http://mathworld.wolfram.com/GraphEdge.html

B. Types of Graph

There a lot of way to represent a graph, we can classify their types by two property. First, graph can be classified based on the existence of parallel edges into two types:

1. Simple Graph
   A simple graph is a graph without a loop or parallel edges connecting a same pair of vertices. In the simple graph, edges does not concerned about the order of the vertices that they connecting to. So, we can say (*u*,*v*) as the same as (*v*,*u*).
2. Multigraph
   Multigraph is a graph that consist a loop or a parallel edges that connect the same pair of vertices. A pseudograph is sub-type of a multigraph that has an edge that connect to the same vertex or there's a loop in the graph.

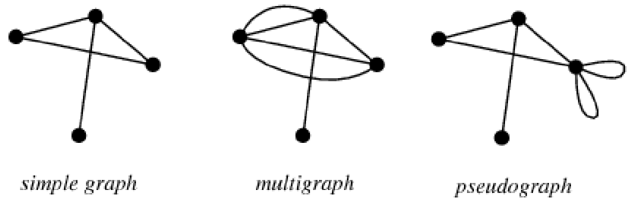

Fig 2. Graph classified by parallel edges of each or a pair of vertices
http://mathworld.wolfram.com/Graph.html

Then, based by their association of vertices pair ordering that connected by the edge or the existence of direction in the edges, the graph has two types:

1. Undirected Graph
   Edges in the undirected graph don't have any direction or we can safely say that no association of ordering the pair of vertices, example (*u*,*v*) would be the same as (*v*,*u*).
2. Directed Graph
   Directed graph is a graph that edges have direction on it or it associated with the order of the pair of vertices, (*u*,*v*) is not the same as (*v*,*u*).

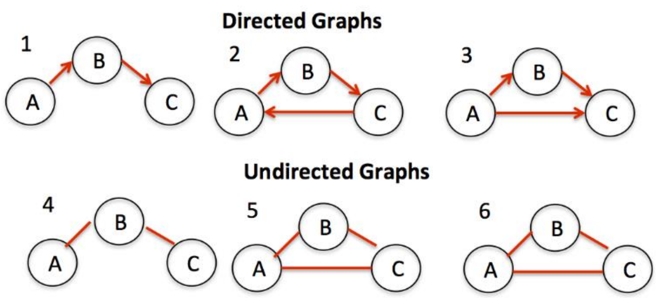

Fig 3. Graph classified by the edges's direction or ordering of the pair of vertices
https://sites.google.com/a/cs.christuniversity.in/discrete-mathematics-lectures/graphs/directed-and-undirected-graph

C. Representing Graphs

We as human can easily analyze an image of a graph by analyze the dots and the lines, but do we ever think how to represent a graph so it can be easier to read by a computer. As we know, the computer read everything in 0's and 1's or we frequently referring as binary number. So by all of these means, We need a systematic and meaningful way to represent graph in other way. There two way two represent a graph:

1. Adjacency Matrix
   Let G = (V, E) be a graph with no multiple edges where V = {1, 2, . . . , n}. The adjacency matrix of G is the n × n matrix A = (aij), where aij = 1 if there is an edge between vertex i and vertex j and aij = 0 otherwise. [4]

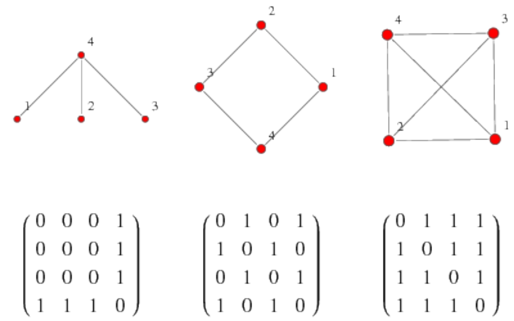

Fig 4. Graph represented with adjacency matrix
http://mathworld.wolfram.com/AdjacencyMatrix.html

2. Incidence Matrix
   Let G = (V, E) be a graph where V = {1, 2, . . . , n} and E = {e1, e2, . . . , em}. The incidence matrix of G is an n × m matrix B = (bik), where each row corresponds to a vertex and each column corresponds to an edge such that if ek is an edge between i and j, then all elements of column k are 0 except bik = bjk = 1. [4]

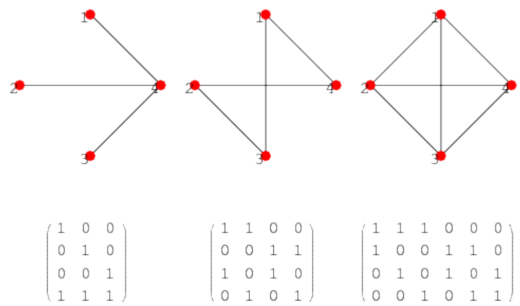

Fig 5. Graph represented with incidence matrix
http://mathworld.wolfram.com/IncidenceMatrix.html

D. Basic Graph-Theoretical Concepts

There are some basic concepts allowing to characterize graphs and real-world networks. We will discuss with the use of terms from graph and network as interchangeable, vertex, site, and node as synonyms, and either edge or link.

- Degree of a Vertex
  The degree *k* of the vertex is the number of edges linking to this node. Nodes having a degree *k*

substantially above the average are denoted “hubs”, they are the VIPs of network theory.[2]

- Coordination Number
  The simplest type of network is the random graph. It is characterized by only two numbers: By the number of vertices *N* and by the average degree *z*, also called the coordination number.[2] The coordination number *z* is the average number of links per vertex. Graph with an average degree *z* has connections with the value of *Nz*/2.
- Erdős – Rényi Random Graphs
  By this, we can construct a specific type of random graph simply by taking *N* nodes or vertices and draw *Nz*/2 amount of lines or edges between randomly choosen pair of nodes.
  For Erdős – Rényi random graphs we have
  $$p = \frac{Nz}{2}\frac{N}{2} = \frac{z}{N-1}$$
  for the relation between the coordination number z and the connection probability p.[2]

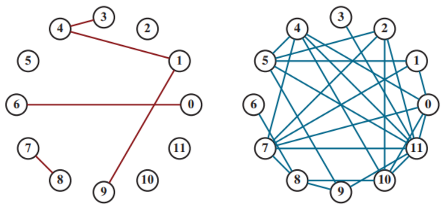

Fig 6. Random graphs with N = 12 vertices and different connection probabilities p = 0.0758 (left) and p = 0.3788 (right). The three mutually connected vertices (0,1,7) contribute to the clustering coefficient and the fully interconnected set of sites (0,4,10,11) is a clique in the network on the right[2]
https://itp.uni-frankfurt.de/~gros/Vorlesungen/CADS/CADS-networks.pdf

- Ensemble Realizations
  There are, for any given link probability p and vertex number N, a large number of ways to distribute the links, compare Fig 4. For Erdős – Rényi graphs every link distribution is realized with equal probability. When one examines the statistical properties of a graphtheoretical model one needs to average over all such “ensemble realizations”.[2]
- The Thermodynamic Limit
  The limit where the number of elements making up a system diverges to infinity is called the “thermodynamic limit” in physics. A quantity is extensive if it is proportional to the number of constituting elements, and intensive if it scales to a constant in the thermodynamic limit.[2]
- Network Diameter and the Small-World Effect
  Network diameter is the maximal separation between pairs of vertices.[2] A random network with *N* amount of vertices and *z* coordination value, we can compute
  $$z\,D \approx N,\ D \propto \log N/\log z$$
  while any node has *z* amount of neighbors, z^2 are the next-nearest neighbors from it and so on. The characteristic of small-world networks is the logarithmic increase in the number of degrees of separation with the network’s size.
- Clustering in Networks
  In the real networks, they have a strong local recurrent connections. The average fraction of pairs of neighbors of a node that are also neighbors of each other is called as the clustering coefficient which labelled *C*. In a random graph a typical site has $z(z-1)/2$ pairs of neighbors. The probability of an edge to be present between a given pair of neighbors is $p = z/(N-1)$. The clustering coefficient, which is just the probability of a pair of neighbors to be interconnected is therefore
  $$Crand = z\,N - 1 \approx z\,N$$
- Cliques and Communities
  The normalized number of triples of fully interconnected vertices is being measured by the clustering coefficient. A clique generally is any fully connected subgraph. A clique is a set of vertices for which every node is connected by every other member of the clique and there is no node outside the clique. The term of “clique” itself came from social networks. A group of friends where everybody knows everybody else is called a clique. In graph theory, a clique corresponds to the maximal fully connected subgraph. In a Erdős – Rényi graphs with *N* amount of vertices and linking probability *p*, the amount of cliques is
  $$\binom{N}{K} p^{K(K-1)/2}(1-p^K)^{N-K}$$
  where the *K* is the size of the clique, $\binom{N}{K}$ is the number of sets of *K* vertices, $p^{K(K-1)/2}$ is the probability of all *K* vertices mutually interconnected, and $(1-p^K)^{N-K}$ is the probability of every *N-K* outside of cliques vertices that not connected to *K* amount vertices from cliques.
- Clustering for Real-World Networks
  The statistical properties of nodes as if they all are independent of each other is being captured by the degree distribution labelled p_k. The property of a given node generally will be dependent however on the other nodes’s properties. While this occurred, it called as “correlation effects” with example of correlation coeffiecient *C*.
- Tree Graphs
  Almost every type of graphs that commonly in graph theory, such as Erdős – Rényi graphs, the thermodinamyc limit vanishes the clustering coeffiecient, and all the loops become irrelavent which make it denoted as a looples graph called “tree graph”.

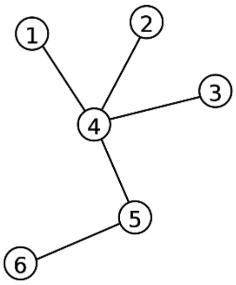

Fig 7. A tree graph
https://en.wikipedia.org/wiki/Tree_(graph_theory)

III. Graph Theory in Six Degrees of Seperation

A. Six Degrees of Seperation

The idea of six degrees of separation grew out of the work carried out in the 1960s by social psychologist Stanley Milgram. Milgram has decided to investigate the so-called question of the small world, the idea that only a few intermediaries bind everyone on the planet. Milgram has decided to investigate the so-called question of the small world, the idea that only a few intermediaries bind everyone on the planet. His project included a hundred people from Boston and Omaha, and he gave them a target in Boston to send a letter from a stranger. The hit is they only can send the letter to a personal friend they thought would be closer to the target. When the test was over, Milgram saw that just six times the letters changed hand. This result famously claimed that we can connect to everyone we wanted to with only six links long of chain of acquaintances.

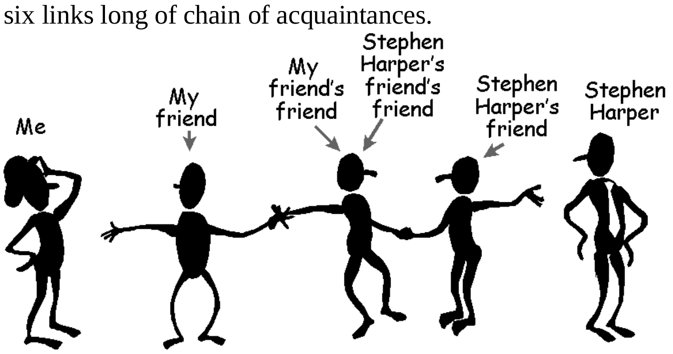

Fig 8. Illustration of six degrees of separation
https://alreadyinexistence.wordpress.com/2013/02/

Think of it: say that I know about 700 people. If each of those people also knows 700 people who are not part of my circle, then in this two-connection network the maximum number of people is (700)(700-1)= 489,300. Now, this number is likely to be inaccurate because with each of my connections I probably have a bunch of mutual friends— our immediate connection networks are highly unlikely to be disconnected. As such, the number of new people I bring to the table will be less than each of my contacts. Still, the principle holds that with each connection, there is a substantial marginal increase in the number of people in the network. [5]

B. Small-World Network

A small-world network is referring to an ensemble network in which mean the shortest path distance between vertices increases sufficiently slowly as function of the number of nodes in network. This term frequently applied to a single network such as family. This term of small-world-network often refered as the Watts–Strogatz toy network.

Small-world networks, according to Watts and Strogatz, are a class of networks that are "highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs." These characteristics result in networks with unique properties of regional specialization with efficient information transfer. Social networks are intuitive examples of this organization, in which cliques or clusters of friends being interconnected but each person is really only five or six people away from anyone else.[7]

The discovery of small-world networks has revolutionized research in network science. In their 1998 landmark paper, Watts and Strogatz described networks that are "highly clustered, like regular lattices, yet have small characteristic path lengths, like random graphs" (Watts and Strogatz, 1998). In other words, small-world networks have the unique ability to have specialized nodes or regions within a network while simultaneously exhibiting shared or distributed processing across all of the communicating nodes within a network.

C. Graph Theory in Small-World Network

- Watts–Strogatz networks

Duncan Watts and Steve Strogatz have formulated the best-known family of small-world networks in a 1998 seminal paper that has helped network science to become a popular medium of expression for many physicists, mathematicians, computer scientists, and many others. Their paper also has a wonderful interactive re-imagination. The word "small-world networks" (or "small-world model") is often used to describe Watts–Strogatz (WS) networks or their variations, although many find it better to identify a more general style of small-worldness. In the literature, however, there is controversy, as others (including the original authors) tend to reserve the term to describe networks with both low mean geodesic path lengths and major local clustering. However, the Watts–Strogatz model predates the analysis of graphs that satisfy the small-world property. For example, contains a proof that the diameter of a network that consists of an N-cycle plus a random matching scales logarithmically with N with probability 1 as $N \to \infty$. (A closed path, which starts and ends at the same node, is called a cycle.) In the matching, one partitions the set of nodes into $N/2$ node pairs if N is even or into $(N-1)/2$ node pairs and one singleton if N is odd; one then adds a new edge between the nodes in each of these pairs.

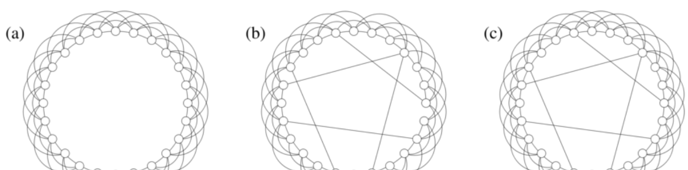

Fig 9. (a) A ring network in which each node is connected to the same number l=3 nearest neighbors on each side. (This network resembles a one-dimensional lattice with periodic

boundary conditions.) (b) A Watts–Strogatz network is created by removing each edge with uniform, independent probability p and rewiring it to yield an edge between a pair of nodes that are chosen uniformly at random. (c) The Newman–Watts variant of a Watts–Strogatz network, in which one adds "shortcut" edges between pairs of nodes in the same way as in a WS network but without removing edges from the underlying lattice.

M. E. J. Newman. The structure and function of complex networks, 2003.

IV. Application Of Six Degrees of Seperation

A. Six Degrees of Kevin Bacon/Oracle of Bacon

The 'Oracle of Bacon' website works on this principle. It uses the IMDB (Internet Movie Database) to make a massive graph of all the actors and movies in the film industry. Actors can be considered nodes, and movies can be considered edges. The website allows users to input any given actor, and see how many edges separate that actor from Kevin Bacon. If a given actor has been in a movie with Kevin Bacon, then his number is 1. If that same given actor has been in a movie with someone else, and that extra person has starred in a film with Kevin Bacon, then the first actor's number is 2. And so on. It seems intuitive that some actors might have pretty high Bacon Numbers, but this is rarely the case. Numbers are astonishingly low. [5]

In this application, we test with a total of fifty sample using Kevin Bacon and relate it with other actors or actresses. Here are the example of the result when we use the application.

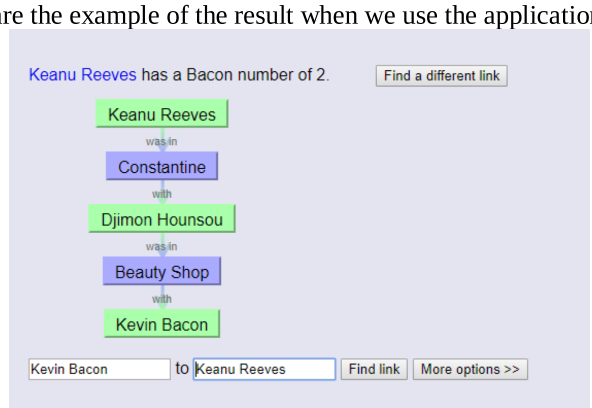

Fig 10. Distribution of degree of shortest path in ArnetMiner https://oracleofbacon.org/movielinks.php

After all of the attempts, we concluded the result that the average Bacon number for an actor/actress to reach for another almost just took three hops. If we analyze, this can be happened because in actors/actresses network, they have a more short and near of clustering. The short the clustering make the less hops it takes to reach another connection that we wanted to.

B. ArnetMiner

ArnetMiner is an online database built by Tsinghua University using semantic web technologies. Currently it contains a society of 0.5M academic researchers. [8]

In this application with ArnetMiner, we analyze by randomly choose a hundred pairs of researchers and search all the path that connecting them. The connections are bounded by the co-authorship of paper or project.

Fig 11. Illustration of collaborator in ArnetMiner Zhang, Lei & Tu, Wanqing., Six Degrees of Separation in Online Society, 2009.

The formed paths are sorted using various metrics. First, we rank all of them based on or defined as the amount of hops by ascending order of the length of the path. It is possible to conclude the shortest length of path as the degree of separation between two researchers. For all 100 pairs, this cycle has been repeated and we can then evaluate the plot from it. The result can be seen below, as it is possible to connect almost over 80 percent with path less than 3 hops that theorized not as much as six degrees of separation.

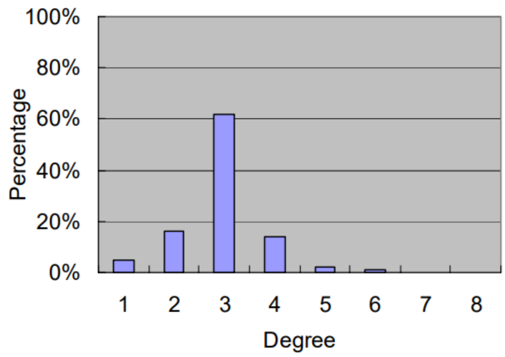

Fig 12. Distribution of degree of shortest path in ArnetMiner Zhang, Lei & Tu, Wanqing., Six Degrees of Separation in Online Society, 2009.

We choose only the shortest path in the first experiment and don't matter if the ties in the path are strong or weak. We related weights in the second set of experiments to show the intensity of the ties. When two authors have more articles and projects co-authored, they will have a stronger connection. Then we figure out all possible paths that connect the two in downward order. As the degree of distance between the two authors, the length of the strongest path is taken. For all the 100 sets, we then repeat this process and plot the distribution curve of them. Yet here is still observed the principle of "Six Degrees of Separation."

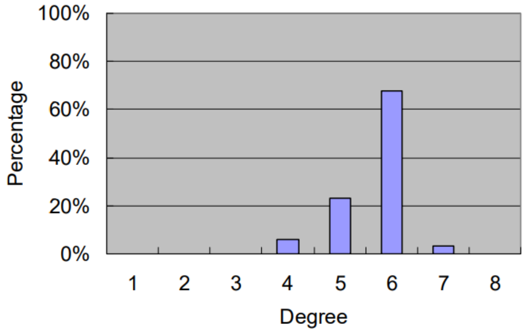

Fig 13. Distribution of degree of strongest path in ArnetMiner Zhang, Lei & Tu, Wanqing., Six Degrees of Separation in Online Society, 2009.

V. Conclusion

The theory of small-world network and six degrees of separation can be proofed, while many factors can disturbed the six degrees of separation. In the case of actors, they have more systematic random graph of connection or the Watts–Strogatz networks. The shorter the distance of clustering from the clique, the less the degree of separation to hops and reach the connection we wanted to.

VI. Acknowledgment

The author wishes to thank God for the capacity granted to the author to finish the article. The author would also like to thank Ms. Fariska Zakhralativa, M.T. as the lecturer of Discrete Mathematics for the guidance given for this semester and other lecturers, Mr. Dr. Ir. Rinaldi Munir, M.T. and Ms. Harlili, M.Sc. for the materials at Discrete Mathematics. For last, the author would also like to thank parents, family, friends, and the open-source application for six degrees of seperation who have supported the author to finish the paper.

References

[1] K. H. Rosen, Discrete Mathematics and Its Applications, 7th Edition, 2013.
[2] Claudius Gros, Complex and Adaptive Dynamical Systems: A Primer, Springer Publishing Company, Incorporated, 2009.
[3] P. S. Dodds, R. Muhamad, D. J. Watts, An experimental study of search in global social networks, 2003.
[4] https://mathcourses.nfshost.com/archived-courses/mat-375-001-2015-fall/lectures/lec-23-adjacency-and-incidence-matrices.pdf degrees at 5 December 2019.
[5] https://blogs.cornell.edu/info2040/2015/10/14/the-small-world-of-kevin-bacon/ accessed at 5 December 2019.
[6] M. Rinaldi, Matematika Diskrit, 3rd Edition, 2006.
[7] Telesford QK, Joyce KE, Hayasaka S, Burdette JH, Laurienti PJ. The ubiquity of small-world networks. Brain Connect, 2011.
[8] Zhang, Lei & Tu, Wanqing., Six Degrees of Separation in Online Society, 2009.

Pernyataan

Dengan ini saya menyatakan bahwa makalah yang saya tulis ini adalah tulisan saya sendiri, bukan saduran, atau terjemahan dari makalah orang lain, dan bukan plagiasi.

Bandung, 6 Desember 2019

Muhammad Daru Darmakusuma – 13518057