Journal of Modern Technology and Engineering
Vol.5, No.3, 2020, pp.255-263

RELEVANT GRAPH CONCEPTS FOR BIG DATA

Aysun Aytaç*

Ege University, Faculty of Science, Department of Mathematics, İzmir, Turkey

Abstract. The subject of much of the creative research and development of graph analytics and visualization has been on data. Graph methods for data structuring, interpretation, and visualization have become critical areas of attention for the application of big data. In order to represent complex interconnections and big groups of closely connected entities, graphs are well adapted. Graphs that model any complex system of a big data domain are so large that they have thousands or even millions of nodes and edges. The main aim of dealing with large graphs is to show the graph properties of a comparatively smaller subgraph. Hence, we will try to provide a solution to big data analytics in this study. So these solutions can be applied to the big data of social networks.

Keywords: Graph theory, big data, graph analytics, centrality.
AMS Subject Classification: 05C90, 68T09, 94A16, 68M10.
Corresponding author: Aysun Aytaç, Ege University, Faculty of Science, Department of Mathematics, İzmir, Turkey, +90 232 311 17 45, e-mail: *aysun.aytac@ege.edu.tr*
Received: 18 April 2020; Accepted: 21 September 2020; Published: 24 December 2020.

1 Introduction

Enormous data networks are everywhere nowadays. Social networks, genetic networks, interactions networks, transportation networks and organizational networks are examples of big data networks (Wasserman & Faust, 1994; Reguly et al., 2006; Merrill, Bakken, Rockoff, Gebbie, & Carley, 2007). The information included in those big data networks deserves efficacious insights and patterns. In addition, these insights can be useful for organizations in taking better business decisions. Hence the big size and the complexity of such networks makes the manual analysis difficult.

Big data is defined as the large amount of data sets that can not be analyzed and managed with traditional data processing tools. What is meant by big data is not only the volumetric size, but also the fact that these data are formed and stored at an increasing speed beside the volume and type, that is, the speed of data generation and data diversity. Files such as social media posts, photos and videos are collected from many sources in different ways. In order to obtain meaningful and valuable information from the large, fast and diverse data collection collected, the data must be made processable. The term big data analysis is used for the methods developed for this purpose. In big data analysis, in addition to data mining, computer science, machine learning, database management, especially mathematical models and algorithms and statistics science come to the fore.

Networking for big data is a cross discipline area. As big data research is in its early stages, the research of big data networking interconnects with the other characteristics of big data, such as big data evolution, mathematical properties, storage distribution, upper layer application expectation and demands. Accordingly, in the global environment of big data, it is vital to analyze networking for big data (Yu, Liu, Dou, Liu, & Zhou, 2017).

Big Data is generally defined as large-scale structural, semi-structural and / or unstructured data that cannot be processed with traditional databases and software (Marz & Warren, 2015). Big data needs new management and processing methods because they contain various types of data, are high-dimensional and are produced at high speeds. While big data was first characterized by the first three features below, in later times, as a result of the change in the definition of big data, five different characteristics began to be characterized (Priya, 2018):

• Volume / Size (Volume): It is the size of the data that needs to be processed. Big data has the gigabyte, which can be processed by traditional systems, has a petabyte or larger data size, which is much larger than the megabyte level.

• Velocity: It indicates the generation frequency and dynamism of the data. In real time applications, it is possible to generate a lot of data in a very short time.

• Variety: It is the emergence of data in various forms, in other words in more than one form. Therefore, the data can occur in different sources.

• Reliability (Veracity): The data is reliable. The data must be able to reach the correct result in analysis. Incomplete data may be encountered in data of high volume and from different sources.

• Value: The data is worth processing. Data that is very valuable for one analysis may be worthless for another analysis.

The big data sets should be represented using graphs intuitively, and then substantial theories and tools for graphs can be applied. Data are considered as graphs in network analysis. Graphs consist of nodes and edges. The attributes correspond to the nodes in the networks and interrelations between several attributes correspond to the edges. Since the interconnections between the nodes of the network are of different kinds, the resulting graph-based structures are frequently very complicated.

The generation of graph theory can be traced back to 1736, and the development of graph theory until the last century, it attracted people's attention in 1940s. Graph theory has very intuitive and straightforward characteristics. In particular when graph theory is applied in order to solve some practical problems, it can be more better to convert the problem into an identical graph theory problem. Due to these properties, graph theory is widely used in modern science. Computer, systems engineering, network engineering, applied mathematics and many other areas are the instances that have been applied and evolved. In simple terms, graph theory is a somewhat older discipline, and the tenacious vitality of graph theory like the human neural network has contributed many to the development of human science and technology (Peretto, 1992).

Graph theory and big data analysis create an effective fusion model. It also makes people recognize that big data analytics and graph theory have specific relations, and can also encourage the study and evolution of big data analytics under the function of graph theory. Big data analytics and graph theory are consolidated. Particularly the solution of difficult problems can be better in the integration of the two (Chen, Wang, & Tian, 2019).

It is worth mentioning that when big data is applied to graph theory problems, the following fundamentals of implementation should be followed. The problem can be solved by modelling it by the use of graph theory. Numerous problems have been solved by related notions of graph theory such as graph coloring, graph covering and so on by the influential fusion of graph theory and big data.

For this reason, in the research of social problems, natural science, engineering technology, and economic management, graph theory is an important modern mathematical tool and has attracted more and more attention in the world of mathematics and other scientific communities. Graphs maintain a robust primitive for modeling data in a diversity of applications. Nodes in graphs in general represent real world objects, and edges correspond to relationships between objects. Social networks, biological networks, and dynamic network traffic graphs are the instances of data modeled as graphs. Graphs are very large, with thousands even millions of nodes and edges in big data applications. By only visual assessment, it is almost impossible to understand the information encoded in large graphs. Graph concepts and graph algorithms are utilized in order to make the graph of big data processable using available tools and techniques (Miller, Ramaswamy, Kochut, & Fard, 2015).

In this paper, we consider simple finite undirected graphs without loops and multiple edges. Let $G = (V, E)$ be a graph with a vertex set $V = V(G)$ and edge set $E = E(G)$. The order of G is the number of vertices in G. The distance $d(u, v)$ between two vertices u and v in G is the length of the shortest path between them. The diameter of G, denoted by $diam(G)$, is the largest distance between two vertices in $V(G)$. The degree $deg_G(v)$ of a vertex $v \in V(G)$ is the number of edges incident to v. For any vertex $v \in V(G)$, the open neighbourhood of v is $N_G(v) = \{u \in V(G) \mid uv \in E(G)\}$ and closed neighbourhood of v is $N_G[v] = N_G(v) \cup \{v\}$. The complement $\bar{G}$ of a graph G has $V(G)$ as its vertex set, but two vertices are adjacent in $\bar{G}$ if only if they are not adjacent in G (Bondy & Murty, 1976).

Social networks have received much attention in recent years since social networks are an important data source of big data applications. Researchers have developed several measures in graph theory to understand and make use of the information in large networks. Centrality is one of these measurements.

Centrality is defined as the relative importance of a node in a graph based on how well the nodes 'connect' the graph. As an example, in a social network, it is important to know how each person is connected with other people. People who are more connected in a graph are more central in the network and thus they are more influential or important in the community represented by that network. Various different types of centrality measures have been proposed over the years (Priya, 2018).

In this study, we only deal with three concepts of centrality. These are vertex betweenness centrality (Unnithan, Kannan, & Jathavedan, 2014; Freeman, 1977; A. Aytac, Ciftci, & Kartal, 2019; A. Aytac & Ciftci, 2019), Edge betweenness centrality (Girvan & Newman, 2002; Boccaletti et al., 2007; Comellas & Gago, 2007; Aytaç & Öztürk, 2018), and closeness centrality (Dangalchev, 2006, 2011; Aytaç & Odabaş, 2011; Aytaç & Odabaş Berberler, 2018). First, vertex betweenness centrality values of wheel graph, complete graph, cycle graph and hypercube graph obtained by Raghavan Unnithan et al. in 2014 are presented. Then, vertex betweenness centrality values for total graphs of some known graphs are computed by Aytac et al. in 2019. The total graph is important since it is the largest graph that is formed by the adjacent relations of nodes and edges of a graph. Second, the concepts of edge betweenness centrality and average edge betweenness centrality are defined. The results obtained by Comellas, F. and Gago, S in 2007 are presented. Then, the computational results are given including the general results for the complement of a graph and the values for some known graph types such as cycle, star, wheel, complete bipartite graph presented by Aytaç and Aksu in 2018. Finally, closeness centrality is defined and the values computed by Dangalchev in 2006 for complete graph, star, path and cycle are presented. Then, the works by Aytac and Berberler in 2011 and 2018 are presented.

2 Relevant Graph Theory Concepts for Big Data

Social networks have gained a lot of attention as a key data source for big data applications in recent years. Researchers have proposed different measures in graph theory to understand and use the data in large networks. Up now, different properties have been introduced, such as its volume, density, average distance, distributions of power-law degrees, clustering coefficient, small-world phenomenon, and centrality. We give definitions of centrality and closeness concepts for big data in this section.

2.1 Vertex betweenness centrality

For a network, there are a number of significant properties. Any of them is that vertices on the shortest path between some vertices (Bader, Kintali, Madduri, & Mihail, 2007; Unnithan et al., 2014). Vertex betweenness centrality is based on the calculation of the shortest path. The significance or centrality of a vertex in a network is defined by it. And also, in the analysis of computer networks, social networks, and many other forms of network data models, it has a key role (Otte & Rousseau, 2002; Latora & Marchiori, 2007; Estrada, 2006; Rubinov & Sporns, 2010). In a communication system, for instance, vertices which have higher centrality value are more significant. Because these vertices pass through more data than the others. As they reside on the greatest number of paths taken by messages, deleting these vertices from the network would cut off communications with others. The centrality of vertex betweenness is thus relevant to the connectivity of a network and its reliability. In this research, in particular, this idea is used in human communication and it suggests that when a person in a group is placed on the shortest communication route joining couples of others, that person is in the central position (Freeman, 1977).

Betweenness centrality $C_B(v)$ for a vertex v is defined as

$$C_B(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

where σ_{st} is the number of shortest paths with vertices s and t as their end vertices, while $\sigma_{st}(v)$ is the number of those shortest paths that include vertex v.
The betweenness centrality of a graph G on n vertices is defined as

$$C_B(G) = \frac{2\sum_{i=1}^{n}\left[C_B\left(v^*\right) - C_B\left(v_i\right)\right]}{(n-1)^2(n-2)}$$

where $C_B\left(v^*\right)$ is the largest value of $C_B\left(v_i\right)$ for any vertex v_i in the given graph G.

Theorem 1. *(Unnithan et al., 2014) The betweenness centrality of a vertex v in a wheel graph W_n, $n > 5$, is given by*

$$C_{\mathrm{B}}(v) = \begin{cases} \frac{(n-1)(n-5)}{2}, & \text{if } v \text{ is central vertex} \\ \frac{1}{2}, & \text{otherwise.} \end{cases}$$

Theorem 2. *(Unnithan et al., 2014) Let K_n be a complete graph on n vertices and $e = (v_i, v_j)$ an edge of it. Then, the betweenness centrality of vertices in $K_n - e$ is given by*

$$C_{\mathrm{B}}(v) = \begin{cases} \frac{1}{n-2}, & \text{if } v \neq v_i, v_j \\ 0, & \text{otherwise.} \end{cases}$$

Theorem 3. *(Unnithan et al., 2014) The betweenness centrality of a vertex in complete bipartite graph $K_{m,n}$ is given by*

$$C_{\mathrm{B}}(v) = \begin{cases} \frac{1}{m}\begin{pmatrix} n \\ 2 \end{pmatrix} & \text{if } \deg(v) = n \\ \frac{1}{n}\begin{pmatrix} m \\ 2 \end{pmatrix} & \text{if } \deg(v) = m. \end{cases}$$

Theorem 4. *(Unnithan et al., 2014) The betweenness centrality of any vertex in a path is the product of the number of vertices on either side of that vertex in the path.*

Theorem 5. *(Unnithan et al., 2014) The betweenness centrality of a vertex v in a cycle C_n is given by*

$$C_{\mathrm{B}}(v) = \begin{cases} \frac{(n-1)(n-3)}{8}, & \text{if } n \text{ is odd} \\ \frac{(n-2)^2}{8}, & \text{if } n \text{ is even.} \end{cases}$$

Theorem 6. *(A. Aytac et al., 2019) Let G be a graph of order n and size m with $\gamma(G) > 2$. For $v \in V(\bar{G})$*

$$\frac{m - \left|\cup_{i=1}^{d} N_G\left[vv_i\right]\right|}{n-3} \leqslant C_{\mathrm{B}}(v) \leqslant m - \left|\cup_{i=1}^{d} N_G\left[vv_i\right]\right|$$

where the vertices $v_1, v_2, \ldots, v_d$ are adjacent vertices of v in G.

Definition 1. *(Behzad, 1967) The total graph $T(G)$ of the graph $G = (V(G), E(G))$ has vertex set $V(G) \cup E(G)$, and two vertices of $T(G)$ are adjacent whenever they are neighbors in G. It is easy to see that $T(G)$ always contains both G and $L(G)$ as induced subgraphs.*

The total graph is the largest graph that is formed by the adjacent relation of elements of a graph. It is important from this respect.

Theorem 7. *(A. Aytac et al., 2019) Let $T\left(K_n\right)$ be total graph of K_n. Then, the betweenness centrality of a vertex v in $T\left(K_n\right)$ is*

$$C_{\mathrm{B}}(v) = \frac{1}{4}(n-1)(n-2).$$

Theorem 8. *(A. Aytac et al., 2019) Let $T\left(K_{1,n}\right)$ be total graph of star and c be central vertex of $K_{1,n}$. Then, the betweenness centrality of a vertex v in $T\left(K_{1,n}\right)$ is*

$$C_{\mathrm{B}}(v) = \begin{cases} n(n-1), & \text{if } v = c \\ 0, & \text{if } v \in V\left(K_{1,n}\right) - \{1\} \\ \frac{n-1}{2}, & \text{if } v \in V(L(K_{1,n})). \end{cases}$$

Theorem 9. *(A. Aytac et al., 2019) Let $T\left(K_{m,n}\right)$ be total graph of $K_{m,n}$ with vertex partition $V_1 \cup V_2 \cup V_3$, where $V_1 = \left\{v = v_i \in V\left(K_{m,n}\right) \mid 1 \leq i \leq m\right\}$, $V_2 = \left\{v = v_j \in V\left(K_{m,n}\right) \mid m+1 \leq j \leq m+n\right\}$, $V_3 = \left\{v = e_{ij} \in V\left(L\left(K_{m,n}\right)\right) \mid 1 \leq i \leq m, m+1 \leq j \leq m+n\right\}$. Then the betweenness centrality of a vertex v in $T\left(K_{m,n}\right)$ is*

$$C_{\mathrm{B}}(v) = \begin{cases} \binom{n}{2}\left(\frac{1}{m}+1\right), & \text{if } v \in V_1 \\ \binom{m}{2}\left(\frac{1}{n}+1\right), & \text{if } v \in V_2 \\ \frac{1}{2}(mn-1), & \text{if } v \in V_3. \end{cases}$$

Theorem 10. *(A. Aytac et al., 2019) Let $T\left(C_n\right)$ be total graph of C_n on $2n$ vertices. Then, the betweenness centrality of a vertex v in $T\left(C_n\right)$ is*

$$C_{\mathrm{B}}(v) = \begin{cases} \frac{(n-1)(n-2)}{4}, & \text{if } n \text{ is odd} \\ \frac{2n^2-3n+2}{8}, & \text{if } n \text{ is even.} \end{cases}$$

Theorem 11. *(A. Aytac et al., 2019) Let $T\left(P_n\right)$ be total graph of P_n with vertex partition $V\left(P_n\right) \cup V\left(L\left(P_n\right)\right)$ where $V\left(P_n\right) = \left\{v = v_i \mid 1 \leq i \leq n\right\}$, $V\left(L\left(P_n\right)\right) = \left\{v = e_{j(j+1)} \mid 1 \leq j \leq n-1\right\}$. Then, the betweenness centrality of a vertex v in $T\left(P_n\right)$*

a) if $v = v_i \in V\left(P_n\right)$

$$C_{\mathrm{B}}(v) = \begin{cases} 2(i-1)(n-i), & 2 \leqslant i \leqslant n-1 \\ 0, & \text{otherwise} \end{cases}$$

b) if $v = e_{j(j+1)} \in V\left(L\left(P_n\right)\right)$

$$C_{\mathrm{B}}(v) = 2(j-1)(n-1-j) + (n-j)\sum_{k=1}^{j-1} \frac{1}{n-k} + j \sum_{k=j+1}^{n-1} \frac{1}{k}.$$

The purpose of betweenness centrality for network analysis is to determine if any vertex is more significant than others. Of course, the significance of the vertex depends on what the graph is reflecting and modeling. For instance, a vertex that has a maximum degree might describe an important and influential individual in dealing with networks that describe links between individuals. However the value of a vertex in a communication network can be determined by the number of the shortest route of which it is a part, as in that case, it can mean its volume of work in relation to the processing and transmission of information. Hence, betweenness centrality is also considered to be a relevant measure for graph and big data analysis. Computing the betweenness centrality of is very complicated as it requires determining the shortest paths in a graph among all couples of vertices. Betweenness centrality's computation for simple graphs is crucial. If a complex network can be split into more petite networks, then the solutions to the problem of optimization on the tinier networks can be applied to solve the problem of optimization on the larger network. In graph operations, betweenness centrality can be studied. It is possible to apply the principles on the vertices examined in this paper to the edges of the graph.

2.2 Edge Betweenness

A graph-theoretic term vertex betweenness was first introduced by Freeman (Freeman, 1977) in the late '70s as a significant element in the reliability, simulation, and measurement analysis of complex networks. Then in 2002, Girvan and Newman (Girvan & Newman, 2002) extended this concept to edges and presented the edge betweenness of an edge as the fraction of the shortest paths that pass along it between pairs of vertices. A specified edge's edge betweenness value is the fraction of the shortest routes that pass through that edge, calculated across all couples of vertices. Both localized and global formation of the graph are considered by this calculation. It is critical to assess the average edge betweenness of many graph types because the average edge betweenness provides knowledge on which edge generates most of the network weakness.

Average edge betweenness of the graph G is defined as $b(G)=\frac{1}{|E|}\sum_{e\in E} b_e$, where $|E|$ is the number of the edges and b_{e} is the edge betweenness of the edge e, defined as $b_e=\sum_{i\neq j} b_e(i,j)$ where $b_e(i,j)=n_{ij}(e)/n_{ij}, n_{ij}(e)$ is the number of geodesics (shortest paths) from vertex i to vertex j that contain the edge e, and n_{ij} is the total number of shortest paths (Boccaletti et al., 2007; Comellas & Gago, 2007).

A complete graph is a simple graph in which every pair of distinct vertices is connected by an edge. The complete graph on n vertices has $n(n-1)/2$ edges. For a complete graph, we have $b\left(G_{\text{complete}}\right)=1$. A path graph is a particularly simple example of a tree, namely on which is not branched at all, that is, contains only vertices of degree two and one. In particular, two of its vertices have degree 1 and all others (if any) have degree 2. For a path graph with n vertices, $|E|=n-1$, and therefore $b\left(G_{\text{path}}\right)=n(n+1)/6$. It is easy to see that $b\left(G_{\text{complete}}\right)\leq b(G)\leq b\left(G_{\text{path}}\right)$.

Lemma 1. *(Comellas & Gago, 2007) Let G be a connected graph and let $e\in E$ be an edge with end vertices $i,j\in V$, then*

a) $b_e(i,j)=1=b_e(j,i)$

b) $2\leq b_e\leq n^2/2$ *if n is even and* $2\leq b_e\leq(n-1)^2/2$ *if n is odd.*

c) $b_e=2(n-1)$ *if one of the end vertices of e has degree 1.*

Theorem 12. *(Aytaç & Öztürk, 2018) Let $\bar{G}$ be the complement graph of G. Then, if G has n vertices, m edges with domination number $\gamma(G)>2$, then the average edge betweenness of $\bar{G}$ is*

$$b(\bar{G})=(n(n-1)+2m)/(n(n-1)-2m).$$

Lemma 2. *(Aytaç & Öztürk, 2018) Label the vertices of C_n as $1, 2, \ldots, n$ and the edges of C_n as $e_1, e_2, \ldots, e_n$. Let $d_{ij}(e_k)$ be the distance between and including the edge e_k. $n_{ij}(e_k)$ is the number of paths which includes the edge e_k with length $d_{ij}(e_k)$ $\quad (1 \le i, j, k \le n$ and $i \ne j)$. The relation between $d_{ij}(e_k)$ and $n_{ij}(e_k)$ in graph C_n is in the following.*

$$\begin{gathered}
\textit{If } d_{ij}(e_k) = 1, \textit{ then } n_{ij}(e_k) = 1 \\
\textit{If } d_{ij}(e_k) = 2, \textit{ then } n_{ij}(e_k) = 2 \\
\textit{If } d_{ij}(e_k) = 3, \textit{ then } n_{ij}(e_k) = 3 \\
\vdots \\
\textit{If } d_{ij}(e_k) = (n-1)/2, \textit{ then } n_{ij}(e_k) = (n-1)/2.
\end{gathered}$$

Theorem 13. *(Aytaç & Öztürk, 2018) If C_n is a cycle, then the average edge betweenness for the cycle C_n with n vertices is*

$$b(C_n) = \begin{cases} \left(n^2 - 1\right)/8, & \textit{if } n \textit{ is odd} \\ n^2/8, & \textit{if } n \textit{ is even.} \end{cases}$$

Theorem 14. *(Aytaç & Öztürk, 2018) If $S_{1,n}$ is a star, then the average edge betweenness for the star $S_{1,n}$ with $n+1$ vertices is $b(S_{1,n}) = n$.*

Theorem 15. *(Aytaç & Öztürk, 2018) If $W_{1,n}$ is a wheel graph, then the average edge betweenness for the wheel graph $W_{1,n}$ $(n \ge 5)$ with $n+1$ vertices is $b(W_{1,n}) = (n-1)/2$.*

Theorem 16. *(Aytaç & Öztürk, 2018) If $K_{m,n}$ is a complete bipartite graph, then the average edge betweenness for the complete bipartite graph $K_{m,n}$ with $m+n$ vertices is $b(K_{m,n}) = \left(m^2 + n^2 - (m+n)\right)/mn + 1$.*

It is crucial to assess average edge betweenness for simple graph class to collect data on which edge is the most susceptible. A specified edge's average edge betweenness value is the fraction of the shortest routes that pass through that edge, calculated across all couples of vertices. Both localized and global formation of the graph are considered by this calculation.

2.3 Closeness Centrality

In order to show how near a vertex is to all other vertices in a network (Priya, 2018), closeness centrality can be utilized. The closeness centrality can be used as a metric of how long information from a single vertex is distributed to all available vertices in the network.

The closeness of a graph is defined as $C = \sum_i C(i)$, where $C(i)$ is the closeness of a vertex i and $C(i) = \sum_{j \ne i} 2^{-d(i,j)}$ (Dangalchev, 2006, 2011; V. Aytac & Turaci, 2018).

Theorem 17. *(Dangalchev, 2006) The closeness of*

a) the complete graph K_n with n vertices is $C(K_n) = (n(n-1))/2$;

b) the star graph S_n with n vertices is $C(S_n) = \frac{(n-1)(n+2)}{4}$;

c) the path P_n with n vertices is $C(P_n) = 2n - 4 + \frac{1}{2^{n-2}}$.

Theorem 18. *(Aytaç & Odabaş, 2011) If C_n is a cycle, then the closeness for the cycle C_n with n vertices is*

$$C(C_n) = \begin{cases} 2n\left(1 - 1/2^{(n-1)/2}\right), & \textit{if } n \textit{ is odd} \\ n\left(2 - 3/2^{n/2}\right), & \textit{if } n \textit{ is even.} \end{cases}$$

Theorem 19. *(Aytaç & Odabaş, 2011) If W_n is a wheel, then the closeness for the wheel W_n with $n+1$ vertices is*

$$C(W_n) = \frac{n(n+5)}{4}.$$

Theorem 20. *(Aytaç & Odabaş Berberler, 2018) For any connected G of order n,* $C(G) \leq n(n-1)/2$.

Theorem 21. *(Aytaç & Odabaş Berberler, 2018) For any connected G of order n,*

$$C(G) \geq n\left(\left(\sum_{i=1}^{\mathrm{diam}(G)-1} 1/2^i\right) + (n - \mathrm{diam}(G))/2^{\mathrm{diam}(G)}\right).$$

Theorem 22. *(Aytaç & Odabaş Berberler, 2018) For any graph G, if diam* $(G) \leq 2$*, then*

$$C(G) = (|V(G)|(|V(G)| - 1) + 2|E(G)|)/4.$$

Corollary 1. *(Aytaç & Odabaş Berberler, 2018) For any graph G, if* $\mathrm{diam}(G) > 3$*, then*

$$C(\bar{G}) = (|V(G)|(|V(G)| - 1) - |E(G)|)/2.$$

3 Conclusion

Big data is the form of converting all data collected from various sources such as social media posts, blogs, photographs, videos, and log files into a meaningful and processable form (Wikipedia, 2019). In this study, the graph theoretical approach of the application of big data in the social network is considered. Identifying important people is vital for many purposes in social networks such as Facebook and Twitter. When the graph model of these large networks is created, it is possible to make big data analysis with some measurements in graph theory. In this study, definitions of measurements such as vertex betweenness centrality, edge betweenness centrality and closeness centrality are given. Also, articles on these measurements, which were previously made by me, are collected. It is thought that the values obtained from the graph structures discussed here are important in terms of giving the researchers an idea for big data analysis in large graphs where a big data environment containing these structures is modeled.

References

Aytac, A., & Ciftci, C. (2019). Betweenness centrality of some complementary prism graphs. *Suleyman Demirel Universitesi Fen Bilimleri Enstitusu Dergisi*, *23*, 277 - 283.

Aytac, A., Ciftci, C., & Kartal, Z. (2019). Betweenness centrality of some total graphs. *Iranian Journal of Science and Technology, Transactions A: Science*, *43*(2), 525–533.

Aytac, V., & Turaci, T. (2018). Closeness centrality in some splitting networks. *Computer Science Journal Of Moldova*, *26*(3), 251-269.

Aytaç, A., & Odabaş, Z. N. (2011). Residual closeness of wheels and related networks. *Internat. J. Found. Comput. Sci.*, *22*(5), 1229–1240. doi: 10.1142/S0129054111008660

Aytaç, A., & Odabaş Berberler, Z. N. (2018). Network robustness and residual closeness. *Rairo Oper. Res.*, *52*(3), 839–847. doi: 10.1051/ro/2016071

Aytaç, A., & Öztürk, H. A. (2018). Average edge betweenness of a graph. *Nonlinear Dyn. Syst. Theory*, *18*(3), 241–252.

Bader, D. A., Kintali, S., Madduri, K., & Mihail, M. (2007). Approximating betweenness centrality. In A. Bonato & F. R. K. Chung (Eds.), *Algorithms and models for the web-graph* (pp. 124–137). Berlin, Heidelberg: Springer Berlin Heidelberg.

Behzad, M. (1967). A criterion for the planarity of the total graph of a graph..

Boccaletti, S., Buldu, J. M., Criado, R., Flores, J., Latora, V., Pello, J., & Romance, M. (2007). Multiscale vulnerability of complex networks. [refereed]. *Chaos*, *17*, 043110. doi: 10.1063/1.2801687

Bondy, J. A., & Murty, U. S. R. (1976). *Graph theory with applications.* American Elsevier Publishing Co., Inc., New York.

Chen, X., Wang, H. H., & Tian, B. (2019). Visualization model of big data based on self-organizing feature map neural network and graphic theory for smart cities. *Cluster Computing*, *22*(6), 13293–13305.

Comellas, F., & Gago, S. (2007). Spectral bounds for the betweenness of a graph. *Linear Algebra Appl.*, *423*(1), 74–80. doi: 10.1016/j.laa.2006.08.027

Dangalchev, C. (2006, June). Residual closeness in networks. *Physica A Statistical Mechanics and its Applications*, *365*, 556-564. doi: 10.1016/j.physa.2005.12.020

Dangalchev, C. (2011). Residual closeness and generalized closeness. *Internat. J. Found. Comput. Sci.*, *22*(8), 1939–1948. doi: 10.1142/S0129054111009136

Estrada, E. (2006, Jan). Virtual identification of essential proteins within the protein interaction network of yeast. *Proteomics*, *6*, 35-40.

Freeman, L. C. (1977, March). A set of measures of centrality based on betweenness. *Sociometry*, *40*(1), 35–41.

Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA*, *99*(12), 7821–7826. doi: 10.1073/pnas.122653799

Latora, V., & Marchiori, M. (2007, jun). A measure of centrality based on network efficiency. *New Journal of Physics*, *9*(6), 188–188. doi: 10.1088/1367-2630/9/6/188

Marz, N., & Warren, J. (2015). *Big data: principles and best practices of scalable real-time data systems.* Manning.

Merrill, J., Bakken, S., Rockoff, M., Gebbie, K., & Carley, K. M. (2007). Description of a method to support public health information management: organizational network analysis. *Journal of biomedical informatics*, *40*(4), 422–428.

Miller, J. A., Ramaswamy, L., Kochut, K. J., & Fard, A. (2015, jul). Research directions for big data graph analytics. In *2015 ieee international congress on big data (bigdata congress)* (p. 785-794). Los Alamitos, CA, USA: IEEE Computer Society. doi: 10.1109/BigDataCongress.2015.132

Otte, E., & Rousseau, R. (2002). Social network analysis: a powerful strategy, also for the information sciences. *Journal of Information Science*, *28*(6), 441-453. doi: 10.1177/016555150202800601

Peretto, P. (1992). *An introduction to the modeling of neural networks.* Cambridge University Press.

Priya, R. (2018). *Retail data analytics using graph database* (Master's thesis). doi: 10.13023/ETD.2018.221

Reguly, T., Breitkreutz, A., Boucher, L., Breitkreutz, B.-J., Hon, G. C., Myers, C. L., ... Tyers, M. (2006). Comprehensive curation and analysis of global interaction networks in saccharomyces cerevisiae. *Journal of Biology*, *5*(4), 11.

Rubinov, M., & Sporns, O. (2010, Sep). Complex network measures of brain connectivity: uses and interpretations. *NeuroImage*, *52*, 1059-69.

Unnithan, S. K. R., Kannan, B., & Jathavedan, M. (2014). Betweenness centrality in some classes of graphs. *International Journal of Combinatorics*, *2014*, 241723.

Wasserman, S., & Faust, K. (1994). *Social network analysis: Methods and applications.* Cambridge University Press. doi: 10.1017/CBO9780511815478

Wikipedia. (2019). *Wikipedia.* Retrieved from `https://en.wikipedia.org/wiki/Big_data`

Yu, S., Liu, M., Dou, W., Liu, X., & Zhou, S. (2017). Networking for big data: A survey. *IEEE Communications Surveys Tutorials*, *19*(1), 531-549. doi: 10.1109/COMST.2016.2610963