JOURNAL OF THE INTERNATIONAL MATHEMATICAL VIRTUAL INSTITUTE
ISSN (p) 2303-4866, ISSN (o) 2303-4947
www.imvibl.org /JOURNALS / JOURNAL
J. Int. Math. Virtual Inst., Vol. **10**(2)(2020), 325-339
DOI: 10.7251/JIMVI2002325A

Former
BULLETIN OF THE SOCIETY OF MATHEMATICIANS BANJA LUKA
ISSN 0354-5792 (o), ISSN 1986-521X (p)

GRAPH THEORY FOR BIG DATA ANALYTICS

Aysun AYTAÇ and Hanife AKSU ÖZTÜRK

Abstract. Big data is a thriving term at the present moment, and it offers researchers numerous opportunities. Big data is a groundbreaking concept in many areas to achieve detailed results and research on human beings' interests. Applications such as the Internet of Things, science research, health care, finance, and e-commerce are applications where tremendous data is generated, and information needs to be obtained properly. Data analysis of big data enables humanity to make more accurate and better decisions for many problems. Analysis of social networks is an implementation of graph theory to explain and identify relationships on social networks. Social media produce vast volumes of data every day that is impossible to manage with conventional data analytics algorithms and methods such as data mining and deep learning. Social network data is helpful for finding interaction between people, analysis of confidence, analysis of effect, the suggestion of any item or place, prediction of connections, identification of crime, etc. In this study, a mathematical graph model of a social network was devised, and edge betweenness centrality, one of the graph-theoretical measurements for social network analytics, was studied.

1. Introduction

Big Data is a term for large data sets having a large, more diverse and complex structure with the difficulties of storing, analyzing, and visualizing for more processes or results [**22**]. Following [**18**], Big data means "data which is too big, too fast, or too hard for existing tools to process." Therefore, it means not only the volumetric size but also means 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 are also emphasized. Big Data has various definitions in the literature, see [**21**].

2010 *Mathematics Subject Classification.* 05C90; 68T09; 94A16; 68M10; 68R10.
Key words and phrases. Graph theory; Big data; Graph analytics; Centrality.
Communicated by Daniel A. Romano.

Social media posts, photos and videos are collected from many sources in diverse ways. In order to obtain meaningful and valuable information from the collected large, fast, and diverse data collections, the data must be processed. The term big data analysis is used for the methods developed for this purpose. In addition to data mining, computer science, machine learning, database management, especially mathematical models and algorithms, and statistics science come to the fore in big data analysis. There are plenty of papers and books on big data [**20, 12, 22, 17, 10**]. Much of big data is available in the form of point clouds in arbitrarily high dimensions. Thus, the science of networks can be of great help. Since networks can be modeled by graphs, this enables us to use the benefits of graph theory.

Graph theory has very intuitive and simple characteristics, which make it widely used in modern science. When solving some practical problems, graph theory can transform the problem into an equivalent graph theory problem. Graphs are significant in computer science due to the ability to abstract a vast class of problems. Graph theory has many application areas for studying and modeling numerous problems such as software plagiarism detection, web search engines, molecular bonds to modeling of social networks like Facebook, LinkedIn and Twitter [**19**].

The management and efficient processing of big data are among the urgent needs to enable research communities to leverage different existing services through interconnected complex networks properly. Modeling of such networks is among the discussed needs to appropriately represent their contents and the interaction between their different components in harmony with their characteristics and constraints. Therefore, there is an essential need to model such networks before focusing on processing and managing big data. Since a graph models complex networks, this helps manage them and question their contents. Thus, graphs have been used to represent data sets, then existing theories and tools for graphs can be applied. Vertices in graphs usually represent real-world objects, and edges show relationships between objects. Some examples of data modeled as graphs are social networks, biological networks, and dynamic network traffic graphs. In big data applications, graphs are very large, having a vast number of vertices and edges. It is almost impossible to understand the information hidden in large graphs by visual inspection alone. To make the big data graph manageable using available tools, relevant graph theoretical concepts and measures such as graph centrality can be used.

For the concept of network-based big data, the concept of decentralized network - based big data structures can be used. According to this concept, the entire network is divided into n multiple logical big data subnetworks [**12**].

The graph we are dealing with here corresponds to any of the n subnetworks described above. In this study, to solve the problems in large graphs, the problem was investigated in known small graphs. In this way, it will provide to make an interpretation for the large graphs containing the small graph structures. This study also aims to determine which vertex or edge in the graph is more important or more robust when a mathematical graph model is obtained for many social networks such as Facebook and Twitter. To do this, there are some measurements defined in graph theory, such as vertex and edge betweenness centrality [**13, 3, 4, 15**], closeness centrality [**6**], residual closeness [**2, 5**]. The field of graph shortest paths has been of significant importance and has wide-spread applications. Thus, in this study, the average edge betweenness centrality measurement [**1**] is discussed in some wheel-related graphs such as friendship, gear, helm, and sunflower. The notion of edge betweenness centrality is based on the number of shortest paths that pass through a certain edge.

In this paper, we consider simple finite undirected graph that has no self-loops and no more than one edge between any two different vertices. Let $G = (V, E)$ be a graph with a vertex set $V = V(G)$ and an edge set $E = E(G)$. The distance (length of a shortest path) between the vertices u_i and u_j of G in denoted by $d(u_i, v_j|G)$. The sum of the distances between all pairs of vertices of G is the distance of the graph G and is denoted by $d(G)$. The distance $d(u, v)$ between two vertices u and v in G is the length of a shortest path between them. If u and v are not connected, then $d(u, v) = \infty$, and for $u = v$, $d(u, v) = 0$. In addition, the distance between the vertices u and v in G can be denoted by $d(u, v|G)$. The diameter of G, denoted by $diam(G)$ is the largest distance between two vertices in $V(G)$ [**9, 8**].

2. Average Edge Betweenness Centrality

Social networks are an important source of data for big data applications and big data analysis. Therefore, social networks have attracted great attention in recent years and today. Researchers have proposed different graph theory approaches to grasp and use the knowledge in large networks.

Centrality is one of these metrics. Centrality, the importance of a vertex in a graph, is defined relative to how well the graph is connected. For example, it can be extremely important in a social network to know to what extent a person is connected with others. In a graph, more linked individuals are more important for that network. Because they are located in the more centers of the network, they have more impact in the social group that this network represents. In the course of the years, various kinds of centrality measure concepts have been introduced in the literature. The average edge betweenness centrality is one of them.

The called the shortest path betweenness, betweenness is defined as the ratio of the number of shortest paths in which a vertex is located to the number of shortest paths that exist between any pair of vertices in the network.

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 [**11, 7, 13, 15**].

Let be compare two graphs G_1 and G_2. If $b(G_1) < b(G_2)$, then G_1 is more stability than G_2. 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 $b(G_{complete}) = 1$. 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:

$$b(G_{path}) = n(n+1)/6.$$

It is easy to see that

$$b(G_{complete}) \leqslant b(G) \leqslant b(G_{path}).$$

The following lemma provides some basic properties for the betweenness related parameters. Let us recall that for a graph G, b_e is the betweenness of edge e, $b(G)$ is the average edge betweenness of G.

Lemma 2.1. [**11**] *Let G be a connected graph and let $e \in E$ be an edge with end vertices $i, j \in V$, then*

i) $b_e(i,j) = 1 = b_e(j,i)$.
ii) $2 \leqslant b_e \leqslant n^2/2$ *if n is even and* $2 \leqslant b_e \leqslant (n-1)^2/2$ *if n is odd.*
iii) $b_e = 2(n-1)$ *if one of the end vertices of e has degree 1.*

Lemma 2.2. [**11**] *Let G be a graph of order n, then*

i) *If e is an edge-bridge of the graph G connecting G_1 with $G - G_1$ where $|V(G_1)| = n_1$, then $b_e = 2n_1(n - n_1)$.*
ii) *If C is a cut-set of edges of the graph G, connecting two sets of vertices X and $V(G) - X$ and $|X| = n_x$, then $\sum\limits_{e\in C} b_e = 2n_x(n - n_x)$.*

3. Average edge betweenness centrality of some graphs

In this section, some general results about the average edge betweennesses centrality measurement and calculated values for some wheel related graphs such as gear, helm, sunflower and friendship graphs are given.

Theorem 3.1. *Let G be a graph with n vertices and m edges. The average edge betweenness of G is*

$$b(G) = d(G)/m.$$

Proof. The edge betweenness of a given edge is the fraction of shortest paths, counted over all pairs of vertices that pass through that edge. The average edge betweenness of a graph G is the fraction of edges, the sum of edge betweenness of all edges. Let i and j be the vertices and e_k is the edge of $G(k = \overline{1,m}, i \neq j)$. It is obvious that

$$\sum_{k=1}^{m} b_{e_k}(i,j) = d(i,j).$$

So the sum of edge betweenness of all edges

$$\sum_{e\in E} b_e = \sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k=1}^{m} = b_{e_k}(i,j) = d(G).$$

Thus

$$b(G) = \sum_{e \in E} b_e/m = d(G)/m.$$

The proof is completed. □

THEOREM 3.2. *If a graph G has n vertices, m edges and $diam(G)$ is diameter of G, then the average edge betweenness of G is*

$$b(G) \leqslant 1 + diam(G)(n(n-1)/2m - 1).$$

PROOF. Let u and v be vertices of G, and let $|E| = m$. So we have m pairs of vertices with $d(u, v|G) = 1$. There are $(n(n-1)/2) - m$ pairs of vertices with $d(u, v|G) \leqslant diam(G)$. Thus the distance of G is

$$d(G) \leqslant m + diam(G)(n(n-1)/2m - 1).$$

Let we divide m both sides of inequality

$$d(G)/m \leqslant (m + diam(G)(n(n-1)/2m - 1))/m.$$

From Theorem 3.1 we have

$$b(G) \leqslant 1 + diam(G)(n(n-1)/2m - 1).$$

The proof is completed. □

COROLLARY 3.1. *Let G be a graph and e is an edge of G. The average edge betweenness of (G-e) is*

$$b(G - e) \geqslant (d(G) + 1)/(m - 1).$$

3.1. Gear Graph (G_n). Gear graph is a wheel graph with a vertex added between each pair adjacent vertices of the outer cycle. Gear graph G_n has $2n + 1$-vertices and $3n$-edges. Gear graph G_n includes an even cycle C_{2n}. The vertices of C_{2n} in G_n are of two kinds: vertices of degree two and three, respectively. The vertices of degree two will be referred to as minor vertices and vertices of degree three to as major vertices [**14**].

Let the central vertex of gear graph G_n be c. The central vertex c has a vertex degree of n. The edges between c central vertex and i the other vertices will be referred to e_{ci} and edges on cycle will be referred to $e_{i(i+1)}$ where $i + 1$ is taken modulo $2n$ ($i = \overline{1, 2n}$).

THEOREM 3.3. *The average edge betweenness of the gear graph G_n for $n \geqslant 3$ is*

$$b(G_n) = 2(n - 1).$$

PROOF. We have five cases for the shortest paths.

Case 1. Let we consider the paths between center vertex c and the major vertices. There is only one path with 1 length e_{ci} ($i = 1, 3, \ldots, 2n - 1$). Thus $b(e_{ci}) = 1$. Hence, for n pairs of vertices, we have

$$\sum_{i=1}^{n} d(c, i) = \sum_{i=1}^{n-1} b_{e_{ci}}(c, i) = n.$$

Case 2. Let we consider the paths between center vertex c and the minor vertices. There are two paths with 2 length $e_{c(i-1)}e_{(i-1)i}$ and $e_{c(i+1)}e_{i(i+1)}$ $(i = 2, 4, \ldots, 2n)$. Thus

$$b(e_{c(i-1)}) = 1/2, b(e_{c(i+1)}) = 1/2, b(e_{(i-1)i}) = 1/2, b(e_{i(i+1)}) = 1/2.$$

Hence, for n pairs of vertices, we have

$$\sum_{i=2}^{2n} b_{e_{c(i-1)}}(c,i) + \sum_{i=2}^{2n} b_{e_{c(i+1)}}(c,i) + \sum_{i=2}^{2n} b_{e_{(i-1)i}}(c,i) + \sum_{i=2}^{2n} b_{e_{i(i+1)}}(c,i) = 2n.$$

Case 3. Let we consider the paths between major vertices. We have two cases for the distance of major vertices between them on the cycle.

Subcase 3.1. If $d(i, i+2) = 2$ on cycle then there are two paths with 2 length between i and $i+2$. These paths are $e_{ci}e_{c(i+2)}$ and $e_{i(i+1)}e_{(i+1)(i+2)}$ where $i = 1, 3, \ldots, 2n-1$, and where $i+1$ and $i+2$ are taken modulo $2n$. Thus

$$b(e_{ci}) = 1/2, b(e_{c(i+2)}) = 1/2, b(e_{i(i+1)}) = 1/2, b(e_{(i+1)(i+2)}) = 1/2.$$

Hence, for n pairs of vertices, we have

$$\sum_{i=1}^{2n-1} b_{e_{ci}}(i,(i+2)) + \sum_{i=1}^{2n-1} b_{e_{c(i+2)}}(i,(i+2)) + \sum_{i=1}^{2n-1} b_{e_{i(i+1)}}(i,(i+2)) + \sum_{i=1}^{2n-1} b_{e_{(i+1)(i+2)}}(i,(i+2)) = 2n.$$

Subcase 3.2. If $d(i,j) > 2$ on cycle then there is only one path with 2 length between i and j. This path is $e_{ci}e_{cj}$ $(i, j = 1, 3, \ldots, 2n-1)$. Thus $b(e_{ci}) = 1$ and $b(e_{cj}) = 1$. Hence, for $n(n-3)/2$ pairs of vertices, we have

$$\sum_{i=1}^{2n-1}\sum_{j=1}^{2n-1} b_{e_{ci}}(i,j) + \sum_{i=1}^{2n-1}\sum_{j=1}^{2n-1} b_{e_{cj}}(i,j) = n(n-3).$$

Case 4. Let we consider the paths between minor vertices. We have three cases for the distance of minor vertices between them on the cycle.

Subcase 4.1. If $d(i, i+2) = 2$ on cycle then there is only one path with 2 length between i and $i+2$. This path is $e_{i(i+1)}e_{(i+1)(i+2)}$ $(i = 2, 4, \ldots, 2n)$, where $i+1$ and $i+2$ are taken modulo $2n$. Thus $b(e_{i(i+1)}) = 1$ and $b(e_{(i+1)(i+2)}) = 1$. Hence, for n pairs of vertices, we have

$$\sum_{i=2}^{2n} b_{e_{i(i+1)}}(i,(i+2)) + \sum_{i=2}^{2n} b_{e_{(i+1)(i+2)}}(i,(i+2)) = 2n.$$

Subcase 4.2. If $d(i, i+4) = 4$ on cycle then there are five paths with 4 length between i and $i+4$. These paths are

$e_{i(i+1)}e_{(i+1)(i+2)}e_{(i+2)(i+3)}e_{(i+3)(i+4)}$; $e_{(i-1)i}e_{c(i-1)}e_{c(i+5)}e_{(i+4)(i+5)}$;

$e_{(i-1)i}e_{c(i-1)}e_{c(i+3)}e_{(i+3)(i+4)}$; $e_{i(i+1)}e_{c(i+1)}e_{c(i+5)}e_{(i+4)(i+5)}$;

$e_{i(i+1)}e_{c(i+1)}e_{c(i+3)}e_{(i+3)(i+4)}$,

where $i \in \{2, 4, , 2n\}$ and $i+1$, $i+2$, $i+3$, $i+4$, $i+5$ are taken modulo $2n$. Thus, we get

$b(e_{i(i+1)}) = 3/5, b(e_{(i+3)(i+4)}) = 3/5, b(e_{(i+1)(i+2)}) = 1/5, b(e_{(i+2)(i+3)}) = 1/5$

$b(e_{(i-1)i}) = 2/5, b(e_{(i+4)(i+5)}) = 2/5, b(e_{c(i-1)}) = 2/5, b(e_{c(i+1)}) = 2/5,$

$b(e_{c(i+3)}) = 2/5, b(e_{c(i+5)}) = 2/5.$

Hence, for n pairs of vertices, we have

$$\sum_{i=2}^{2n} b_{e_{i(i+1)}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{(i+3)(i+4)}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{(i+1)(i+2)}}(i,(i+4))$$
$$+ \sum_{i=2}^{2n} b_{e_{(i+2)(i+3)}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{(i-1)i}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{(i+4)(i+5)}}(i,(i+4))$$
$$+ \sum_{i=2}^{2n} b_{e_{c(i-1)}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{c(i+1)}}(i,(i+4))$$
$$+ \sum_{i=2}^{2n} b_{e_{c(i+3)}}(i,(i+4)) + \sum_{i=2}^{2n} b_{e_{c(i+5)}}(i,(i+4)) = 4n.$$

Subcase 4.3. If $d(i,j) > 4$ on cycle then there are four paths with 4 length between i and j. These paths are

$e_{(i-1)i}e_{c(i-1)}e_{c(j+1)}e_{j(j+1)}$;

$e_{(i-1)i}e_{c(i-1)}e_{c(j-1)}e_{(j-1)j}$;

$e_{i(i+1)}e_{c(i+1)}e_{c(j+1)}e_{j(j+1)}$;

$e_{i(i+1)}e_{c(i+1)}e_{c(j-1)}e_{(j-1)j}$,

where $i, j \in \{2, 4, \dots 2n\}$ and $i+1$, $j+1$ are taken modulo $2n$. Thus, we get

$b(e_{(i-1)i}) = 2/4, b(e_{j(j+1)}) = 2/4, b(e_{i(i+1)}) = 2/4, b(e_{(j-1)j}) = 2/4,$

$b(e_{c(i-1)}) = 2/4, b(e_{c(j-1)}) = 2/4, b(e_{c(i+1)}) = 2/4, b(e_{c(j+1)}) = 2/4.$

Hence, for $n(n-5)/2$ pairs of vertices, we have

$$\sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{(i-1)i}}(i,j) + \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{j(j+1)}}(i,j) + \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{i(i+1)}}(i,j) +$$
$$\sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{(j-1)j}}(i,j) + \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{c(i-1)}}(i,j) + \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{c(j-1)}}(i,j)$$
$$+ \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{c(i+1)}}(i,j) + \sum_{i=2}^{2n}\sum_{j=2}^{2n} b_{e_{c(j+1)}}(i,j) = 2n(n-5).$$

Case 5. Let we consider the paths between minor and major vertices. We have three cases for the distance of minor and major vertices between them on the cycle.

Subcase 5.1. If $d(i,i+1)=1$ on cycle then there is only one path with 1 length between i and $i+1$. This path is $e_{i(i+1)}$, where $i=\overline{1,2n}$ and $i+1$ is taken modulo $2n$. Thus $b(e_{i(i+1)})=1$. Hence, for $2n$ pairs of vertices, we have

$$\sum_{i=1}^{2n} b_{e_{i(i+1)}}(i,(i+1))=2n.$$

Subcase 5.2. If $d(i,i+3)=3$ on cycle then there are three paths with 3 length between i and $i+3$. These paths are

$e_{i(i+1)}e_{(i+1)(i+2)}e_{(i+2)(i+3)}$; $e_{ci}e_{c(i+2)}e_{(i+2)(i+3)}$; $e_{ci}e_{c(i+4)}e_{(i+3)(i+4)}$,

where $i=\overline{1,2n}$ and $i+1$, $i+2$, $i+3$, $i+4$ are taken modulo $2n$. Thus we get

$b(e_{i(i+1)})=1/3$, $b(e_{(i+3)(i+4)})=1/3$, $b(e_{(i+1)(i+2)})=1/3$,

$b(e_{ci})=2/3$, $b(e_{c(i+2)})=1/3$ and $b(e_{c(i+4)})=1/3$.

[3pt]Hence, for $2n$ pairs of vertices, we have

$$\sum_{i=1}^{2n} b_{e_{i(i+1)}}(i,(i+3))+\sum_{i=1}^{2n} b_{e_{(i+3)(i+4)}}(i,(i+3))+\sum_{i=1}^{2n} b_{e_{(i+1)(i+2)}}(i,(i+3))$$
$$+\sum_{i=1}^{2n} b_{e_{(i+2)(i+3)}}(i,(i+3))+\sum_{i=1}^{2n} b_{e_{(i+3)(i+4)}}(i,(i+3))+\sum_{i=1}^{2n} b_{e_{ci}}(i,(i+3))$$
$$+\sum_{i=1}^{2n} b_{e_{c(i+2)}}(i,(i+3))+\sum_{i=1}^{2n} b_{e_{c(i+4)}}(i,(i+3))=6n.$$

Subcase 5.3. If $d(i,j)>3$ on cycle then there are two paths with 3 length between i and j. These paths are $e_{ci}e_{c(j+1)}e_{j(j+1)}$; $e_{ci}e_{c(j-1)}e_{(j-1)j}$, where $i=\overline{1,2n}$ and $j+1$ are taken modulo $2n$. Thus, we get

$$b(e_{c(j+1)})=1/2,\ b(e_{j(j+1)})=1/2,\ b(e_{ci})=2/2,\ b(e_{c(j-1)})=1/2$$

and $b(e_{(j-1)j})=1/2$. Hence, for $n(n-4)$ pairs of vertices, we have

$$\sum_{i=1}^{2n}\sum_{j=1}^{2n} b_{e_{c(j+1)}}(i,j)+\sum_{i=1}^{2n}\sum_{j=1}^{2n} b_{e_{j(j+1)}}(i,j)+\sum_{i=1}^{2n}\sum_{j=1}^{2n} b_{e_{ci}}(i,j)$$
$$+\sum_{i=1}^{2n}\sum_{j=1}^{2n} b_{e_{c(j-1)}}(i,j)+\sum_{i=1}^{2n}\sum_{j=1}^{2n} b_{e_{(j-1)j}}(i,j)=3n(n-4).$$

As a consequence by Case 1 - Case 5, the average edge betweenness of G_n is

$$b(G_n) = (1/3n)(n+2n+2n+n(n-3)+2n+4n+2n(n-5)+2n+6n+3n(n-4)) = 2(n-1).$$

The proof is completed. □

3.2. Helm Graph (H_n). Helm H_n is a graph obtained from a wheel W_n with cycle C_n having a pendant edge attached to each vertex of the cycle. Helm graph H_n consists of the vertex set

$$V(H_n) = \{i|1 \leqslant i \leqslant n\} \cup \{i'|1 \leqslant i \leqslant n\} \cup \{c\}$$

and the edge set

$$E(H_n) = \{e_{i(i+1)}|1 \leqslant i \leqslant n\} \cup \{e_{ii'}|1 \leqslant i \leqslant n\} \cup \{e_{ci}|1 \leqslant i \leqslant n\}.$$

The vertices of $H_n - \{c\}$ are of two kinds: vertices of degree four and one, respectively. The vertices of degree one will be referred to as minor denoted by i' and vertices of degree four to as major vertices denoted by i [**14, 16**].

THEOREM 3.4. *The average edge betweenness of the helm graph H_n for $n \geqslant 3$ is*

$$b(H_n) = 2(n-1).$$

PROOF. We have four cases for the shortest paths.

Case 1. Let we consider the paths between center vertex c and the major vertices. Also we consider the paths between major vertices. It is the same as wheel graph W_n so we have ($i, j = \overline{1,n}$ and $i \neq j$)

$$\sum_{i=1}^{n} b_{e_{ci}} + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ij}} = n(n-1).$$

Case 2. Let we consider the paths between center vertex c and the minor vertices. There is only one path with 2 length $e_{ci}e_{ii'}$, ($i = \overline{1,n}$). Thus $b(e_{ci}) = 1$ and $b(e_{ii'}) = 1$. Hence, for n pairs of vertices we have

$$\sum_{i=1}^{n} b_{e_{ci}}(c,i) + \sum_{i=1}^{n} b_{e_{ii'}}(c,i) = 2n.$$

Case 3. Let we consider the paths between minor vertices. We have three cases for the distance of major vertices between them on the cycle. Let i and j be the major vertices of H_n ($i = \overline{1,n}$, $i \neq j$ and $i < j$).

Subcase 3.1. If $d(i,j) = 1$ on cycle then there is one path between i' and j'. This path is $e_{ii'}e_{ij}e_{jj'}$. Thus, we have $b(e_{ii'}) = 1$, $b(e_{ij}) = 1$ and $b(e_{jj'}) = 1$. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ii'}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ij}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i',j') = 3n.$$

Subcase 3.2. If $d(i,j) = 2$ on cycle then there are two paths between i' and j'. These paths are $e_{ii'}e_{i(i+1)}e_{(i+1)j}e_{jj'}$ and $e_{ii'}e_{ci}e_{cj}e_{jj'}$. Thus, $b(e_{i(i+1)}) = 1/2$,

$b(e_{(i+1)j}) = 1/2$, $b(e_{ci}) = 1/2$, $b(e_{cj}) = 1/2$, $b(e_{ii'}) = 1$ and $b(e_{jj'}) = 1$ are obtained. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{i(i+1)}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)j}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ii'}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i',j') = 4n.$$

Subcase 3.3. If $d(i,j) > 2$ on cycle then there is one path between i' and j'. This path is $e_{ii'}e_{ci}e_{cj}e_{jj'}$. Thus, we have $b(e_{ii'}) = 1$, $b(e_{ci}) = 1$, $b(e_{cj}) = 1$ and $b(e_{jj'}) = 1$. Hence, for $n(n-5)/2$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ii'}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i',j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i',j') = 2n(n-5).$$

Case 4. Let we consider the paths between minor and major vertices. Let i be a major vertex and j' be a minor vertex $(i,j = \overline{1,n})$. $d(i,j')$ is the distance between i and j' vertices on cycle. We have four cases.

Subcase 4.1. If $d(i,j') = 1$ then there is one path between i and j'. This path is $e_{ii'}$ and $b(e_{ii'}) = 1$. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{n} b_{e_{ii'}}(i,j') = n.$$

Subcase 4.2. If $d(i,j') = 2$ then there is one path between i and j'. This path is $e_{ij}e_{jj'}$ and we get $b(e_{ij}) = 1$, $b(e_{jj'}) = 1$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ij}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i,j') = 2(2n) = 4n.$$

Subcase 4.3. If $d(i,j') = 3$ then there are two paths between i and j'. These paths are $e_{i(i+1)}e_{(i+1)j}e_{jj'}$ and $e_{ci}e_{cj}e_{jj'}$. Thus we have $b(e_{jj'}) = 1$, $b(e_{i(i+1)}) = 1/2$, $b(e_{(i+1)j}) = 1/2$, $b(e_{ci}) = 1/2$ and $b(e_{cj}) = 1/2$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{i(i+1)}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)j}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i,j') = 3(2n) = 6n.$$

Subcase 4.4. If $d(i,j') \geqslant 4$ then there is one path between i and j'. This path is $e_{ci}e_{cj}e_{jj'}$. Thus we have $b(e_{jj'}) = 1$, $b(e_{ci}) = 1$ and $b(e_{cj}) = 1$.
Hence, for $n(n-5)$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i,j') = 3n(n-5).$$

As a consequence by Case 1 - Case 4 the average edge betweenness of H_n is
$b(H_n) =$
$(1/3n)(n(n-1) + 2n + 3n + 4n + 2n(2n-5) + n + 4n + 6n + 3n(n-5))$
$= 2(n-1)$.
The proof is completed. □

3.3. Sunflower graph (SF_n). Sunflower graph SF_n is defined as follows: consider a wheel with central vertex c and an n-cycle $1, 2, \dots, n$ and additional n vertices $1', 2', \dots, n'$, where i' is joined by i, $(i+1)$ for $i = 1, 2, \dots, n$ where $(i+1)$ is taken modulo n. SF_n has order $2n+1$ and size $4n$. The vertices of $SF_n - \{c\}$ are two kinds: vertices of degree five and two, respectively. The vertices of degree two will be referred to as minor vertices and vertices of degree five to as major vertices. The edges between c and i can be labeled as e_{ci} . The edges on cycle can be labeled as $e_{i(i+1)}$ where $(i+1)$ is taken modulo n and the edges between i major and j' minor vertices can be labeled as $e_{ij'}$ [**14, 16**].

THEOREM 3.5. *The average edge betweenness of the sunflower graph SF_n for $n \geqslant 5$ is*

$$b(SF_n) = (3n-5)/2.$$

PROOF. We have four cases for the shortest paths.

Case 1. Let we consider the paths between center vertex c and the major vertices. Also we consider the paths between major vertices. It is the same as wheel graph W_n so we have ($i, j = \overline{1, n}$ and $i \neq j$)

$$\sum_{i=1}^{n} b_{e_{ci}} + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ij}} = n(n-1).$$

Case 2. Let we consider the paths between center vertex c and the minor vertex i'. There are two paths with 2 length $e_{ci}e_{ii'}$, and $e_{c(i+1)}e_{(i+1)i'}$, ($i = \overline{1, n}$ where $i+1$ is taken modulo n). Thus $b(e_{ci}) = 1/2$, $b(e_{ii'}) = 1/2$, $b(e_{c(i+1)}) = 1/2$ and $b(e_{(i+1)i'}) = 1/2$. Hence, for n pairs of vertices we have

$$\sum_{i=1}^{n} b_{e_{ci}}(c,i') + \sum_{i=1}^{n} b_{e_{ii'}}(c,i') + \sum_{i=1}^{n} b_{e_{c(i+1)}}(c,i') + \sum_{i=1}^{n} b_{e_{(i+1)i'}}(c,i') = 2n.$$

Case 3. Let we consider the paths between minor vertices. We have four cases for the distance of minor vertices between them on the cycle. Let i' and j' be the minor vertices of SF_n ($i, j = \overline{1, n}$, where j, $i+1$ are taken modulo n).

Subcase 3.1. If $d(i', j') = 2$ on cycle then there is one path between i' and j'. This path is $e_{ji'}e_{jj'}$. Thus $b(e_{ji'}) = 1$ and $b(e_{jj'}) = 1$. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ji'}}(i', j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i', j') = 2n.$$

Subcase 3.2. If $d(i', j') = 3$ on cycle then there is one path between i' and j'. This path is $e_{(i+1)i'}e_{(i+1)j}e_{jj'}$. Thus $b(e_{(i+1)i'}) = 1$, $b(e_{(i+1)j}) = 1$ and $b(e_{jj'}) = 1$. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)i'}}(i', j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)j}}(i', j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i', j') = 3n.$$

Subcase 3.3. If $d(i', j') = 4$ on cycle then there are five paths between i' and j' from Subcase 4.2 of Case 4 in Theorem 3.3. and also there is one path on cycle with four length. Hence, for n pairs of vertices, we obtain $4n$.

Subcase 3.4. If $d(i', j') > 4$ on cycle then there are four paths with four length between i' and j' from Subcase 4. 2 of Case 4 in Theorem 3.3. Hence, for $n(n-7)/2$ pairs of vertices, we obtain the total value $4n(n-7)/2 = 2n(n-7)$.

Case 4. Let we consider the paths between minor and major vertices. Let i be a major vertex and j' be a minor vertex $(i, j = \overline{1, n})$. $d(i, j')$ is the distance between i and j' vertices on cycle. We have four cases.

Subcase 4.1. If $d(i, j') = 1$ then there are two pairs of vertices. One path is $e_{ii'}$ and the other one is $e_{(i+1)i'}$ so $b(e_{ii'}) = 1$ and $b(e_{(i+1)i'}) = 1$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ii'}}(i, j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)i'}}(i, j') = 2n.$$

Subcase 4.2. If $d(i, j') = 2$ then there is one path between i and j'. This path is $e_{ij}e_{jj'}$ and $b(e_{ij}) = 1$, $b(e_{jj'}) = 1$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ij}}(i, j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i, j') = 2(2n) = 4n.$$

Subcase 4.3. If $d(i, j') = 3$ then there are two paths between i and j'. These paths are $e_{i(i+1)}e_{(i+1)j}e_{jj'}$ and $e_{ci}e_{cj}e_{jj'}$. Thus we have $b(e_{jj'}) = 1$,$b(e_{i(i+1)}) = 1/2$, $b(e_{(i+1)j}) = 1/2$, $b(e_{ci}) = 1/2$ and $b(e_{cj}) = 1/2$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{i(i+1)}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{(i+1)j}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i,j') = 3(2n) = 6n.$$

Subcase 4.4. If $d(i,j') \geqslant 4$ then there is one path between i and j'. This path is $e_{ci}e_{cj}e_{jj'}$. Thus we have $b(e_{jj'}) = 1$, $b(e_{ci}) = 1$ and $b(e_{cj}) = 1$.
Hence, for $n(n-6)$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i,j') + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{jj'}}(i,j') = 3n(n-6).$$

As a consequence by Case 1 - Case 4 the average edge betweenness of SF_n is

$$b(SF_n) = (1/4n)(n(n-1)+2n+2n+3n+4n+2n(n-7)+2n+4n+6n+3n(n-6)) = (3n-5)/2.$$

The proof is completed.

□

3.4. Friendship graph (f_n). Friendship graph f_n is collection of n triangles with a common point. The friendship graph f_n has two kinds of vertex. It has $2n$ vertices of degree 2 called minor vertex and 1 vertex of degree $2n$ called center vertex. Label the minor vertices as $1, 2, ..., 2n$ and the center vertex as c. Label the edges which are between c and i as e_{ci} $(i = \overline{1, 2n})$ and which are between i and $i+1$ (i is odd) as $e_{i(i+1)}$ [**14, 16**].

THEOREM 3.6. *The average edge betweenness of the friendship graph f_n is*

$$b(f_n) = 1 + 4(n-1)/3.$$

PROOF. We have three cases in order to find the shortest paths.

Case 1. Let we consider the paths between center vertex c and minor vertices i. There is one path e_{ci} between c and i. Thus $b(e_{ci}) = 1$. Hence, for $2n$ pairs of vertices, we obtain

$$\sum_{i=1}^{2n} b_{e_{ci}}(c,i) = 2n.$$

Case 2. Let we consider the paths between minor vertex i and minor vertex $(i+1)$ (i is odd). There is one path $e_{i(i+1)}$ between i and $(i+1)$. Thus $b(e_{i(i+1)}) = 1$. Hence, for n pairs of vertices, we obtain

$$\sum_{i=1}^{2n-1} b_{e_{i(i+1)}}(i,i+1) = n.$$

Case 3. Let we consider the paths between minor vertex i and minor vertex j. There is one path $e_{ci}e_{cj}$ between i and j ($1 \leqslant i,\, j \leqslant 2n$ and if $i(j)$ is odd then $j \neq i+1$ ($i \neq j+1$)). Thus $b(e_{ci}) = 1$ and $b(e_{cj}) = 1$. Hence, for $2n(2n-2)/2$ pairs of vertices, we obtain

$$\sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{ci}}(i,j) + \sum_{i=1}^{n}\sum_{j=1}^{n} b_{e_{cj}}(i,j) = 2n(2n-2).$$

As a consequence by Case 1 - Case 3 the average edge betweenness of f_n is

$$b(f_n) = (1/3n)(2n + n + 2n(2n-2)) = 1 + 4(n-1)/3.$$

The proof is completed. □

4. Conclusion

In this paper we consider the concept of average edge betweenness of some G graphs because when computing $b(G)$, we can gather information on which edge carries the most of the graph vulnerability. The average edge betweenness value is calculated for four kinds of wheel related graphs: gear, helm, sunflower and friendship. Let we put in order to these values as follow:

$$b(f_n) < b(SF_n) < b(G_n) = b(H_n)(n > 13).$$

We see that the f_n graph is more vulnerable then the others. Therefore, for using f_n graph or the graphs which have f_n as a subgraph is more suitable for network design.

References

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

[2] A. Aytaç and Z. N. Odabaş Berberler. Network robustness and residual closeness. *RAIRO Oper. Res.*, **52**(3)(2018), 839–847.

[3] A. Aytaç, C. Çiftçi and Z. Kartal. Betweenness centrality of some total graphs. *Iran. J. Sci. Technol. Trans. A Sci.*, **43**(2)(2019), 525–533.

[4] A. Aytaç and C. Çiftçi. Betweenness centrality of some complementary prism graphs. *Süleyman Demirel Üniversitesi Fen Bilimleri Enstitüsü Dergisi*, **23**(2)(2019), 277–283.

[5] A. Aytaç and Z. N. Odabaş. Residual closeness of wheels and related networks. *Int. J. Found. Comput. Sci.*. **22**(5)(2011), 1229–1240

[6] A. Aytaç and T. Turacı. Closeness centrality in some splitting networks. *Comput. Sci. J. Moldova*, **26**(3)(2018), 251–269.

[7] S. Boccaletti, J. M. Buldú, R. Criado, J. Flores, V. Latora, J. Pello and M. Romance. Multiscale vulnerability of complex networks. *Chaos*, **17**(4)(2007), 43110.

[8] J. A. Bondy and U. S. R. Murty. *Graph theory with applications.* American Elsevier Publishing Co., Inc., New York, 1976.

[9] G. Chartrand and L. Lesniak. *Graphs and Digraphs.* California Wadsworth and Brooks, 2004.

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

[11] F. Comellas and S. Gago. Spectral bounds for the betweenness of a graph. *Linear Algebra Appl.* , **423**(1)(2007), 74-80.
[12] M. Dehmer, F. Emmert-Streib, S. Pickl and A. Holzinger. (Eds.) *Big data of complex networks*. Chapman & Hall/CRC Big Data Series. CRC Press, Boca Raton, FL, 2017.
[13] L. C. Freeman. A set of measures of centrality based on betweenness. *Sociometry*, **40**(1)(1977), 35–41.
[14] J. A. Gallian. *A dynamic survey of graph labeling*. Electron. J. Combin., 15, Dynamic Survey 6. https://www.combinatorics.org/files/Surveys/ds6/ds6v6-2001.pdf
[15] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci. USA* , **99**(12)(2002), 7821–7826.
[16] I. Javaid and S. Shokat. On the partition dimension of some wheel related graphs. *J. Prime Res. Math.*, **4**(2008), 154–164.
[17] M. I. Jordan and T. M. Mitchell. Machine learning: Trends, perspectives, and prospects. *Science*, **349**(6245)(2015), 255–260.
[18] S. Madden. From databases to big data. *IEEE Internet Computing*, **16**(3)(2012), 4–6.
[19] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel and B. Bhattacharjee. Measurement and analysis of online social networks. In Proceedings of the 7th ACM SIGCOMM conference on Internet measurement (pp. 29-42). 2007. https://doi.org/10.1145/1298306.1298311
[20] S. Yu, M. Liu, W. Dou, X. Liu and S. Zhou. Networking for big data: A survey. *IEEE Communications Surveys & Tutorials*, **19**(1)(2016), 531–549.
[21] J. S. Ward and A. Barker. Undefined by data: a survey of big data definitions. arXiv preprint arXiv:1309.5821.
[22] X. Wu, X. Zhu, G. Q. Wu and W. Ding. Data mining with big data. *IEEE transactions on knowledge and data engineering*, **26**(1)(2013), 97–107.

Received by editors 07.12.2020; Available online 14.12.2020.

Aysun AYTAÇ,
Department of Mathematics, Faculty of Science, Ege University, 35100 Izmir, Turkey
E-mail address: aysun.aytac@ege.edu.tr

Hanife AKSU ÖZTÜRK,
Department of Mathematics, Faculty of Science, Ege University, 35100 Izmir, Turkey
E-mail address: hanifeaksu@hotmail.com