AMSI VACATIONRESEARCH SCHOLARSHIPS 2022–23

SET YOUR SIGHTS ON RESEARCH THIS SUMMER

An investigation into properties of the closeness centrality of a graph

Yuqi Liu

Supervised by Thomas Britz

University of New South Wales

20 February 2023

AMSI AUSTRALIAN MATHEMATICAL SCIENCES INSTITUTE

Prelude

Abstract

This project investigated the closeness centrality of the star graphs using both mathematical and computational methods. In this report, the closeness centrality of the equal path length star graph has been found and verified using Wolfram Mathematica. Additionally, the formula has been used to generalise further the closeness centrality of the multi-length path star graph.

Acknowledgements

I would like to thank Thomas Britz for his delightful guidance and ongoing support throughout the busy period this project was done. I have been very privileged to have Thomas as my mentor and also as my discrete math lecturer, thank you for your enthusiasm towards graph theory which got me interested in this project.

Statement of Authorship

Thomas Britz was the primary supervisor for this report, assisted in drawing graph using Tikz, provided feedback and proofread this report.

The definitions mentioned in this report were adapted from other journal articles, and references have been provided where necessary.

Everything else is the work of the author, including proving of the formula, writing of the report and the Wolfram Mathematica code.

Contents

1 Introduction

Closeness centrality is one of the many centrality measures used in social network analysis, including identifying the most influential person in a social network, key infrastructure nodes in the urban networks and super-spreaders of disease [1].

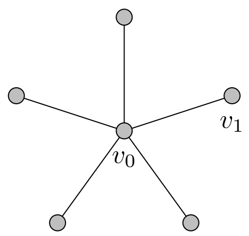
Closeness centrality was defined in 1950 by Alex Bavelas [2]. Since then, however, only few properties of this centrality measure have been established. In this AMSI project, we investigated the star graph and generalised the closenesss centrality formula for it.

Let $G = (V, E)$ be a connected graph on $|V| = n$ vertices. The closeness centrality measure $\bar{\mathsf{C}}_C(G)$ of G is defined as the average of the closeness centralities $\bar{\mathsf{C}}_C(v)$ of each vertex v in G [2]:

$$\bar{\mathsf{C}}_C(G) := \frac{1}{n}\sum_{v\in V}\bar{\mathsf{C}}_C(v) \quad \text{where} \quad \bar{\mathsf{C}}_C(v) := \frac{n-1}{\sum_{w\in V} d(v,w)}$$

and where $d(v, w)$ is the distance between v and w; that is, the shortest length of a path between v and w in G.

Example 1. *Let S_n be a star graph with central vertex v_0 and five outer vertices.*



star graph S_n

$$\bar{\mathsf{C}}_C(v_0) = \frac{n-1}{\sum_{w\in V} d(v_0,w)} = \frac{6-1}{1+1+1+1+1} = 1$$

$$\bar{\mathsf{C}}_C(v_1) = \frac{n-1}{\sum_{w\in V} d(v_1,w)} = \frac{6-1}{1+2+2+2+2} = \frac{5}{9}$$

$$\bar{\mathsf{C}}_C(S_n) = \frac{1}{n}\sum_{v\in V}\bar{\mathsf{C}}_C(v) = \frac{1}{6}\left(1+\frac{5}{9}\times 4\right) = \frac{29}{54}$$

2 Known results on closeness centralities for graph families

There seem to be very few papers on closeness centralities in the mathematical research literature. One of these few papers, by Hu, Islam and Britz [4], provides the closeness centralities for twelve families of graphs, as described in the following two propositions.

Proposition 1. *The vertex closeness centralities $\bar{\mathsf{C}}_C(v)$ for all $v \in V$ are given below for the families of graphs $G = (V, E)$ shown in Figure 1.*

Complete graph K_n: $\bar{\mathsf{C}}_C(v) = 1$

Cycle graph C_n: $\bar{\mathsf{C}}_C(v) = \frac{n-1}{\lfloor n^2/4 \rfloor}$

Wheel graph W_n: $\bar{\mathsf{C}}_C(v) = 1$ *if v is the central vertex, and* $\bar{\mathsf{C}}_C(v) = \frac{n-1}{2n-5}$ *otherwise*

Star graph S_n: $\bar{\mathsf{C}}_C(v) = 1$ *if v is the central vertex, and* $\bar{\mathsf{C}}_C(v) = \frac{n}{2n-1}$ *otherwise*

Near-complete graph $K_n - e$: $\bar{\mathsf{C}}_C(v) = \frac{n-1}{n}$ *if v is adjacent to e, and* $\bar{\mathsf{C}}_C(v) = 1$ *otherwise*

Cocktail party graph $CP(n)$: $\bar{\mathsf{C}}_C(v) = \frac{2n-1}{2n}$

Complete bipartite graph $K_{m,k}$: $\bar{\mathsf{C}}_C(v) = \begin{cases} \frac{m+k-1}{m+2k-2} & \text{if } v \in \{u_1, \ldots, u_k\} \\ \frac{m+k-1}{k+2m-2} & \text{if } v \in \{v_1, \ldots, v_m\} \end{cases}$

Crown graph S_n^0: $\bar{\mathsf{C}}_C(v) = \frac{2n-1}{3n}$

Path graph P_n: $\bar{\mathsf{C}}_C(v_k) = \frac{4(n-1)}{(2k-n+1)^2+n^2-1}$ *for* $k = 0, \ldots, n-1$

Ladder graph L_n: $\bar{\mathsf{C}}_C(v_k) = \bar{\mathsf{C}}_C(u_k) = \frac{4n-2}{(2k-n+1)^2+n^2+2n-1}$ *for* $k = 0, \ldots, n-1$

Circular ladder graph CL_n: $\bar{\mathsf{C}}_C(v) = \frac{n-1}{2\lfloor n^2/4 \rfloor + n}$

Hypercube graph Q_k: $\bar{\mathsf{C}}_C(v) = \frac{2^k-1}{k2^{k-1}}$.

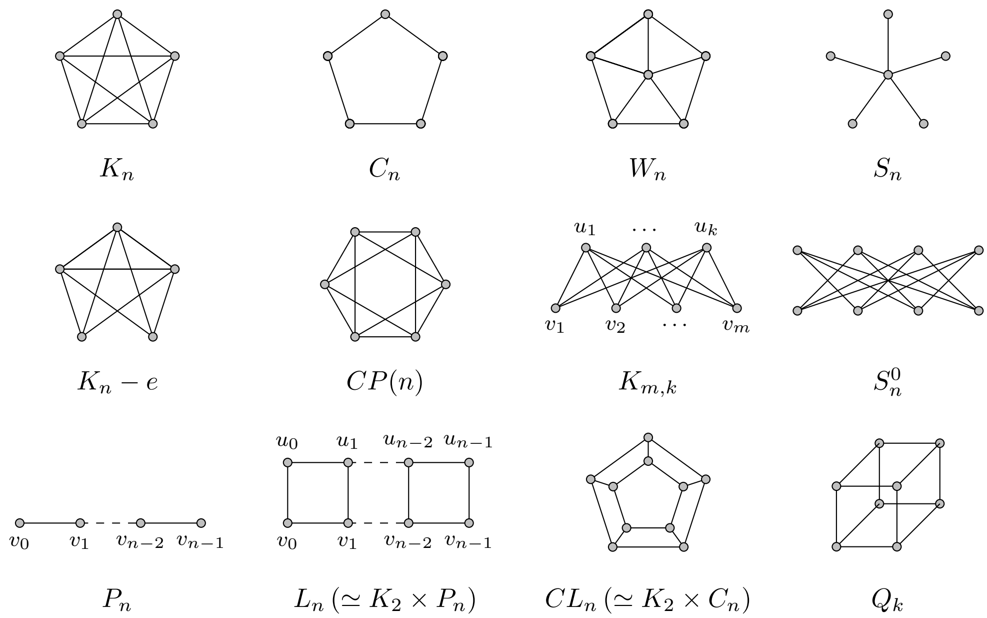


Figure 1: Twelve families of graphs

The expressions for $\bar{\mathsf{C}}_C(v)$ in Proposition 1 were used by Hu, Islam and Britz [4] to find simple expressions for the closeness centrality $\bar{\mathsf{C}}_C(G)$ of ten of the twelve families of graphs G.

Corollary 2. [4]

$$\begin{aligned}
\bar{\mathsf{C}}_C(K_n) &= 1 & \bar{\mathsf{C}}_C(CP_n) &= \tfrac{2n-1}{2n}\\
\bar{\mathsf{C}}_C(C_n) &= \tfrac{n-1}{\lfloor n^2/4\rfloor} & \bar{\mathsf{C}}_C(K_{m,k}) &= \tfrac{m+k-1}{m+k}\left(\tfrac{m}{k+2m-2}+\tfrac{k}{m+2k-2}\right)\\
\bar{\mathsf{C}}_C(W_n) &= \tfrac{n^2-4}{n(2n-5)} & \bar{\mathsf{C}}_C(S_n^0) &= \tfrac{2n-1}{3n}\\
\bar{\mathsf{C}}_C(S_n) &= \tfrac{n^2+2n-1}{2n^2+n-1} & \bar{\mathsf{C}}_C(CL_n) &= \tfrac{n-1}{2\lfloor n^2/4\rfloor+n}\\
\bar{\mathsf{C}}_C(K_n-e) &= \tfrac{n^2-2}{n^2} & \bar{\mathsf{C}}_C(Q_k) &= \tfrac{2^k-1}{k2^{k-1}}\,.
\end{aligned}$$

Less simple expressions were also obtained for the graphs P_n and L_n.

Corollary 3. [4] *For the graphs* $G \in \{P_n, L_n\}$,

$$\bar{\mathsf{C}}_C(G) = -\sum_{a,b=\pm1} \frac{abc}{n\hat{n}}\psi\left(\frac{1+an+b\hat{n}}{2}\right)$$

where $\phi(x)$ *is the digamma function and where* $\hat{n} = \sqrt{1-n^2}$ *and* $c = \frac{1}{n}$ *if* $G = P_n$, *and* $\hat{n} = \sqrt{1-2n-n^2}$ *and* $c = \frac{2n-1}{4n}$ *if* $G = L_n$.

3 The closeness centrality of generalised star graphs

The main results of this report are new expressions for the closeness centrality of a general family of graphs not previously determined in the literature.

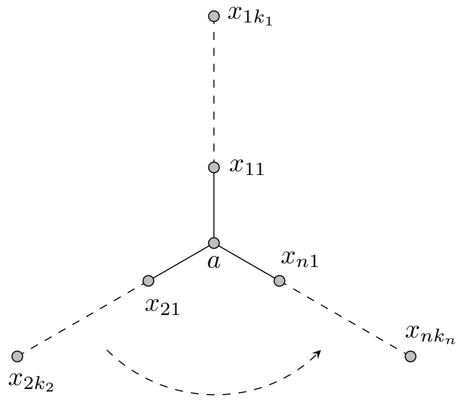


Figure 2: Multi-length path star graph G

Proposition 4. *Let* G *be a multi-length path star graph with vertices* x_{ij}, *where* $1 \le i \le n$ *and* $1 \le j \le k_i$ *for some positive integers* $n, k_1, \ldots, k_n$; *see Figure 2. The vertex closeness centrality for each vertex* x_{ij} *is:*

$$\bar{\mathsf{C}}_C(x_{ij}) = \frac{\displaystyle\sum_{r=1}^{n} k_r}{\displaystyle\sum_{r=1}^{n} \frac{k_r(k_r+1+2j)}{2} + j(j-2k_i)}\,.$$

Proof. Note that

$$\bar{\mathsf{C}}_c(x_{ij}) = \frac{n-1}{\sum\limits_{w\in V} d(v,w)} = \frac{\sum\limits_{i=1}^{n} k_i}{\sum\limits_{r\neq i}\sum\limits_{s=1}^{k_r}(j+s) + \sum\limits_{r=0}^{k_i}|r-j|}$$

$$= \frac{\sum\limits_{i=1}^{n} k_r}{\sum\limits_{r\neq i}\sum\limits_{s=1}^{k_r} j + \sum\limits_{r\neq i}\sum\limits_{s=1}^{k_r} s + \sum\limits_{r=0}^{j}(j-r) + \sum\limits_{r=j+1}^{k_i}(r-j)}$$

Now,

$$\begin{aligned}
&\sum_{r\neq i}\sum_{s=1}^{k_r} j + \sum_{r\neq i}\sum_{s=1}^{k_r} s + \sum_{r=0}^{j}(j-r) + \sum_{r=j+1}^{k_i}(r-j)\\
&= \sum_{r\neq i} k_i j + \sum_{r\neq i}\frac{k_r(k_r+1)}{2} + \frac{1}{2}j(j+1) + \frac{1}{2}(k_i-j)(k_i-j+1)\\
&= j(\sum_{r=1}^{n} k_i) - jk_i + \sum_{r=1}^{n}\frac{k_r(k_r+1)}{2} - \frac{k_i(k_i+1)}{2} + j^2 + \frac{k_i}{2} - jk_i + \frac{{k_i}^2}{2}\\
&= j(\sum_{r=1}^{n} k_r) + \sum_{r=1}^{n}\frac{k_r(k_r+1)}{2} + j^2 - 2jk_i\\
&= \sum_{r=1}^{n}\frac{k_r(k_r+1+2j)}{2} + j(j-2k_i)\,.
\end{aligned}$$

Therefore,

$$\bar{\mathsf{C}}_c(x_{ij}) = \frac{\sum\limits_{r=1}^{n} k_r}{\sum\limits_{r=1}^{n}\frac{k_r(k_r+1+2j)}{2} + j(j-2k_i)}\,.$$

□

By Proposition 4, the vertex closeness centrality for the central vertex a can be found.

$$\bar{\mathsf{C}}_c(a) = \frac{n-1}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{k_i} j} = \frac{2\sum\limits_{r=1}^{n} k_r}{\sum\limits_{i=1}^{n} k_i(k_i+1)}\,.$$

Corollary 5. *Let G be a multi-length path star graph with vertices x_{ij}, where $1 \leq i \leq n$ and $1 \leq j \leq k_i$ for some positive integers $n, k_1, \ldots, k_n$; see Figure 2. The graph closeness centrality for G is*

$$\bar{\mathsf{C}}_C(G) = \frac{\sum\limits_{r=1}^{n} k_r}{1+\sum\limits_{r=1}^{n} k_r}\left(\sum_{i=1}^{n}\sum_{j=1}^{k_i}\frac{1}{j(j-2k_i)) + \frac{1}{2}\sum\limits_{r=1}^{n} k_r(k_r+1+2j)} + \frac{2}{\sum\limits_{i=1}^{n} k_i(k_i+1)}\right).$$

3.1 Checking the correctness of the propositions

The proposition regarding the vertex closeness centrality established above is complicated and may be reused in future projects. To prevent errors, Wolfram Mathematica was used as a computaional tool to ensure the correctness of the propositions.

Let G_1 be a rooted tree with four branches with $7, 3, 4, 6$ edges respectively. Additionally, G_1 is a tree formed by the union of paths

$$P_1 = ax_{11}, x_{11}x_{12}, \dots, x_{16}x_{17};$$

$$P_2 = ax_{21}, x_{21}x_{22}, x_{22}x_{23};$$

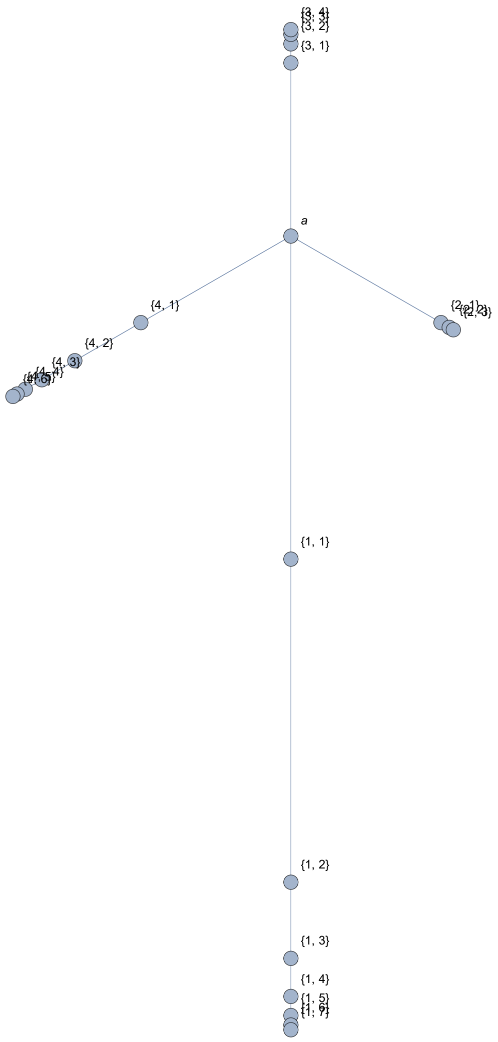
$$P_3 = ax_{31}, x_{31}x_{32}, \dots, x_{33}x_{34};$$

$$P_4 = ax_{41}, x_{41}x_{42}, \dots, x_{45}x_{46}\,.$$

The graph G_1 is represented and drawn in Mathematica as follows.

```
In[1]:= Clear["Global`*"];
        a1 = 7,3,4,6;
        a2 = Flatten[Table[Table[{i, j} [UndirectedEdge] {i, j + 1},
                                 {j, a1[i] - 1}], {i, Length[a1]}]];
        a3 = Table[a [UndirectedEdge] {i, 1}, {i, Length[a1]}];
        a4 = Join[a2, a3];
        a5 = Graph[a4, VertexLabels -> "Name",
        GraphLayout -> {"BalloonEmbedding", "EvenAngle" -> True, "OptimalOrder" -> True}]
```

Out[1]=



In the graph above, the distance between $x_{i_1 j_1}$ and $x_{i_2 j_2}$ is $j_1 + j_2$. The distance between a and x_{ij} is j.

To test the correctness of the program, two simple example were given. The below command used the built-in function 'GraphDistance' in Mathematica to calculate the distance between the node $\{1, 3\}$ and $\{2, 2\}$.

```
In[2]:= GraphDistance[a5,{1,3},{2,2}]

Out[2]= 5
```

Similarly, the following command was used to calculate the distance between the central vertex a and $\{1, 3\}$.

```
In[3]:= GraphDistance[a5,a,1,3,]

Out[3]= 3
```

The vertex closeness centrality $\bar{\mathsf{C}}_C(x_{1,3})$ can be found using the 'GraphDistance' function.

```
In[4]:= a6[v_]:=GraphDistance[a5,v];
        a7[v_] := (Length[a6[v]] - 1) / Apply[Plus, a6[v]]; a7[{1, 3}]
```

Out[4]= $\frac{5}{23}$

On the other hand, the vertex closeness centrality $\bar{C}_C(x_{1,3})$ can also be found by constructing the formula found in Proposition 4 in Mathematica,

$$\bar{C}_C(x_{ij}) = \frac{\sum_{r=1}^{n} k_r}{\sum_{r=1}^{n} \frac{k_r(k_r+1+2j)}{2} + j(j-2k_i)}$$

where $i = 1$ and $j = 3$, as shown in the code below:

```
In[5]:= ii, jj = {1, 3};
        Apply[Plus, a1] / (Apply[Plus, Map[# (# + 1 + 2 jj) / 2 &, a1]] + jj (jj - 2 a1[ii]))
```

Out[5]= $\frac{5}{23}$

Both methods provide the same result, $\bar{C}_C(x_{1,3}) = \frac{5}{23}$, so the formula in Proposition 4 is proven to be correct.

3.2 The closeness centrality for equal-length path star graphs

This subsection presents the closeness centrality for equal-length path star graphs.

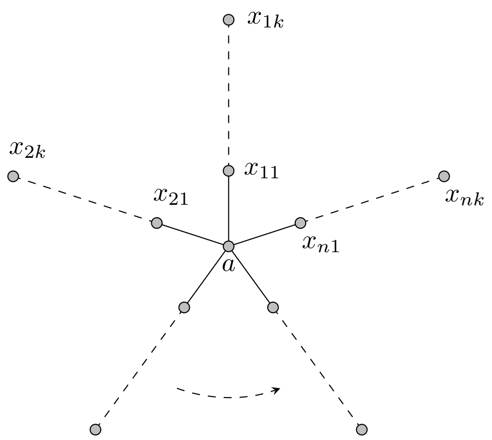


Figure 3: equal-length path star graph G_2

Corollary 6. *Let G_2 be an equal-length path star graph with n paths each of length k, for some positive integers k and n. The vertex closeness centrality for each vertex x_{ij} is*

$$\bar{\mathsf{C}}_C(x_{ij}) = \frac{\sum_{r=1}^{n} k}{\frac{1}{2}\sum_{r=1}^{n} k(k+1+2j)+j(j-2k)}.$$

Proof. From Proposition 4, we have

$$\bar{\mathsf{C}}_C(x_{ij}) = \frac{\sum_{r=1}^{n} k_r}{\frac{1}{2}\sum_{r=1}^{n} k_r(k_r+1+2j)+j(j-2k_i)} = \frac{\sum_{r=1}^{n} k}{\frac{1}{2}\sum_{r=1}^{n} k(k+1+2j)+j(j-2k)}.$$

□

Corollary 7. *Let G_2 be an equal-length path star graph with n paths each of length k, for some positive integers k and n. The graph closeness centrality for G_2 is*

$$\bar{\mathsf{C}}_C(G_2) = \frac{2n^2k}{1+nk}\sum_{j=1}^{k}\frac{1}{2j(j-2k)+nk(k+1+2j)} + \frac{2}{(1+nk)(k+1)}.$$

Proof. From Corollary 5, the graph closeness centrality for multi-length path star graph is

$$
\begin{aligned}
\bar{C}_C(G_2) &= \frac{\sum_{r=1}^{n} k_r}{1+\sum_{r=1}^{n} k_r}\Big(\sum_{i=1}^{n}\sum_{j=1}^{k_i}\frac{1}{j(j-2k_i))+\frac{1}{2}\sum_{r=1}^{n} k_r(k_r+1+2j)}+\frac{2}{\sum_{i=1}^{n} k_i(k_i+1)}\Big)\\
&= \frac{nk}{1+nk}\Big(n\sum_{j=1}^{k}\frac{1}{j(j-2k)+\frac{n}{2}k(k+1+2j)}+\frac{2}{nk(k+1)}\Big)\\
&= \frac{2n^2k}{1+nk}\Big(\sum_{j=1}^{k}\frac{1}{2j(j-2k)+nk(k+1+2j)}\Big)+\frac{2}{(1+nk)(k+1)}
\end{aligned}
$$

□

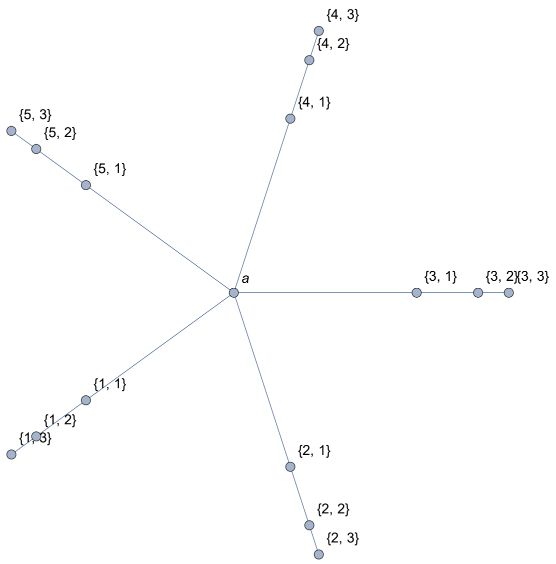
3.3 Checking the correctness of the graph closeness centrality

The corollary regarding the vertex closeness centrality established in Section 5.2 can be further verified via applied mathematical tools. To prevent errors, Wolfram Mathematica was used to ensure the correctness of Corollary 4.

Let G_2 be a rooted tree with five branches each with equal length of three edges. The graph G_2 is defined and drawn in Mathematics as follows.

```
In[6]:= Clear["Global`*"];
        a1 = ConstantArray[3, 5];
        a2 = Flatten[Table[Table[{i, j} [UndirectedEdge] {i, j + 1},
                                  {j, a1[[i]] - 1}], {i, Length[a1]}]];
        a3 = Table[a [UndirectedEdge] {i, 1}, {i, Length[a1]}];
        a4 = Join[a2, a3];
        a5 = Graph[a4, VertexLabels -> "Name",
        GraphLayout ->{"BalloonEmbedding", "EvenAngle" -> True, "OptimalOrder" -> True}]
```

Out[6]=



Similarly, we could use the built-in function 'GraphDistance' and 'VertexList' in Mathematica to calculate $\bar{C}_C(G_2)$ graphically.

```
In[7]:= a6[v_] := GraphDistance[a5, v];
        a7[v_] := (Length[a6[v]] - 1)/Apply[Plus, a6[v]];
        Mean[Map[a7, VertexList[a5]]]
```

$$\text{Out[7]=}\ \frac{5667}{18304}$$

On the other hand, we could also calculate $\bar{C}_C(G_2)$ by substituting $n = 5$ and $k = 3$ into the formula in Corollary 7.

$$\text{In[8]:=}\ f[n\_, k\_] := \frac{2n\hat{}2k}{1+nk}\sum_{j=1}^{k}\frac{1}{2j(j-2k)+nk(k+1+2j)}+\frac{2}{(1+nk)(k+1)}$$

```
        f[5, 3]
```

$$\text{Out[8]=}\ \frac{5667}{18304}$$

Both methods provided the same output $\bar{C}_C(G_2) = \frac{5667}{18304}$ which validates Corollary 7.

4 Discussion and conclusion

This report has provided the explicit expressions for the vertex and graph closeness centrality for multi-length path star graphs, as well as the expressions for equal-length path star graphs.

Additionally, computational methods have been used to further verify the correctness of the expression.

There are many interesting graphs' closeness centrality properties that still remain undiscovered and further investigation is worth looking into in the future.

References

[1] Wikipedia, *Centrality*, `https://en.wikipedia.org/wiki/Centrality`, last viewed on 19 February 2023.

[2] A. Bavelas, Communication patterns in task oriented groups, *J. Acoust. Soc. Amer.* **57** (1950), 271–282.

[3] S. Zhu et al., The connectedness of the coronavirus disease pandemic in the world: A study based on complex network analysis, *Frontiers in Physics* **8** (2021), 14 pages.

[4] X. Hu, A. Islam and T. Britz, Bounds on the closeness centrality of a graph, arXiv:2204.11283 (2022), 9 pages.