ISSN 0005-1179 (print), ISSN 1608-3032 (online), Automation and Remote Control, 2025, Vol. 86, No. 5, pp. 457–463.

OPTIMIZATION, SYSTEM ANALYSIS, AND OPERATIONS RESEARCH

Relations between Average Shortest Path Length and Another Centralities in Graphs

M. A. Tuzhilin[*,**]

[*] *National Research University "Higher School of Economics", Moscow, Russia*
[**] *Lomonosov Moscow State University, Moscow, Russia*
e-mail: avarnavsky@hse.ru

Received December 16, 2024
Revised December 30, 2025
Accepted January 9, 2025

Abstract—Relations between average shortest path length and radiality, closeness, stress centralities and average clustering coefficient were obtained for simple connected graphs.

Keywords: networks, centralities, local and global properties of graphs, average shortest path length, Watts-Strogatz clustering coefficient

DOI: 10.31857/S0005117925050078

1. INTRODUCTION

The average shortest path length is one of the most important characteristics of real graphs (or networks). The first essential attempt for calculation this value was in 1969 year by two social psychologists Stanley Milgram and Jeffrey Travers (small-world experiment [1]). They found a phenomenon that any two people know each other by a sequence of 6 contacts. This theory is known as six degrees of separation theory. Further, this theory was proved by Duncan Watts and Steven Strogatz in [2]. It turns out, that many graphs in real applied works have the same property: they have small average shortest path length.

Another very important characteristic for real networks was found by Watts and Strogatz — big average clustering coefficient (or Watts and Strogatz clustering coefficient [2]). Graphs which satisfied these properties (small average shortest path length and big average clustering coefficient) are called small-world networks. In Social Networks science there are many other important characteristics for networks so-called centralities (for example, stress, closeness, betweenness, radiality centralities and so on). In this field they analyzed distributions of all these centralities to understand intrinsic characteristics of networks (see for example [3–7]), but the existence of relations between centralities is still not very well-known question. In these article we found relations between average shortest path length and radiality, closeness, stress centralities and also the connection between average shortest path length and average clustering coefficient for geodetic graphs.

2. MAIN DEFINITIONS

All subsequent definitions are given for a simple, connected, undirected graph G. Let's denote by

- $V(G)$ the set of vertices, $E(G)$ – the set of edges, $A = \{a_{ij}\}$ adjacency matrix of graph G,
- neighborhood $N(v)$ – the induced subgraph in G on vertices which adjacent to the vertex v,
- $N'(v)$ induced subgraph in G on vertices $= V\big(N(v)\big) \bigcup \{v\}$,

- $\bar{f}(x_1, \ldots, x_k)$ for any function $f : V \times V \times \ldots \times V \to \mathbb{R}$ – the restriction of this function on $N'(v)$ (for example $\bar{L}(x, y)$ will be the average shortest path between x and y with restriction to subgraph $N'(v)$),
- $d_i = deg(v_i)$,
- $\mathrm{dist}(s, t)$ – the shortest path length between s and t,
- $n = |V(G)|, \quad m = |E(G)|$,
- $X(i) = X(v_i)$ for any X – set or function corresponding to vertex v_i.

Let's give definitions of centralities.

(1) Diameter $\mathrm{diam}(G) = \max_{s,t \in V(G)} \mathrm{dist}(s, t)$.

(2) Average shortest path length
$L(G) = \frac{1}{n(n-1)} \sum\limits_{s,t \in V(G),\ s \neq t} \mathrm{dist}(s, t)$.

(3) Local cluster coefficient
$c_i = c(i) = \frac{\text{number of edges in } N(i)}{\text{maximum possible number of edges in } N(i)} = \frac{2|E(N(i))|}{d_i(d_i - 1)}$.

(4) Average clustering coefficient

$$C_{WS}(G) = \frac{1}{n} \sum_{i \in V(G)} c_i = \frac{1}{n} \sum_{i \in V(G)} \frac{2|E(N(i))|}{d_i(d_i - 1)} = \frac{1}{n} \sum_{i \in V(G)} \frac{\sum\limits_{j,k \in V(G)} a_{ij} a_{jk} a_{ki}}{d_i(d_i - 1)}.$$

(5) Global clustering coefficient

$$C(G) = \frac{\text{number of closed triplets in } G}{\text{number of all triplets in } G} = \frac{\sum\limits_{i,j,k \in V(G)} a_{ij} a_{jk} a_{ki}}{\sum\limits_{i \in V(G)} d_i(d_i - 1)}.$$

(6) Closeness centrality $\mathrm{Clo}(v) = \frac{n-1}{\sum\limits_{t \in V(G)} \mathrm{dist}(v,t)}$.

(7) Radiality $\mathrm{Rad}(v) = \frac{\sum\limits_{t \in V(G), t \neq v} (\mathrm{diam}(G) + 1 - \mathrm{dist}(v,t))}{n-1}$.

(8) Stress $\mathrm{Str}(i) = \sum\limits_{s,t \in V(G),\ s \neq t \neq i} \sigma_{st}(i)$, where $\sigma_{st}(i)$ – is the total number of shortest paths from s to t which contains vertex i.

Note that all centralities are non-negative and $c_i, C_{WS}, \mathrm{Clo}(v)$ are less or equal 1.

Also let's give the definition of geodetic graph:

Definition 1. If there exists the unique shortest path between any two vertices in G then the graph G is called **geodetic.**

This definition is equivalent to the condition then there is no even cycles in a graph.

3. MAIN RESULTS

Let's consider a induced subgraph $G' \subset G$. In general G' can be not connected graph. In this case let's define the average shortest path length for vertices of G' with relation to the distance dist in the ambient graph G. Let's call $L(N(i))$ the local average shortest path length for the vertex i.

Let's start with simple relations: let's proof the relation between local shortest path length and local clustering coefficient.

Lemma 1.

$$L(N(i)) = 2 - c_i.$$

Proof.

$$L(N(i)) = \frac{1}{d_i(d_i-1)} \sum_{s,t \in N(i), s \neq t} \operatorname{dist}(s,t)$$

$$= \frac{1}{d_i(d_i-1)} \sum_{(s,t) \in E(N(i))} \operatorname{dist}(s,t) + \sum_{s,t \in N(i), (s,t) \notin E(N(i))} \operatorname{dist}(s,t)$$

$$= \frac{1}{d_i(d_i-1)} \left(2|E(N(i))| + \sum_{(s,i),(i,t) \in E(G), (s,t) \notin E(G)} \operatorname{dist}(s,t) \right)$$

$$= \frac{1}{d_i(d_i-1)} \Big(2|E(N(i))| + 2(d_i(d_i-1) - 2|E(N(i))|) \Big) = 2 - c_i.$$

Note that shortest paths for vertices in $N(i)$ are defined corresponding to whole graph G.

Averaging by i we obtain simple corollary about the relation between local shortest path length and average clustering coefficient.

Corollary 1.

$$C_{WS}(G) = 2 - \frac{1}{n} \sum_{i \in V(G)} L(N(i)).$$

Let's proof the relation between shortest path length in subgraph $N'(i)$ and shortest path length in $N(i)$.

Lemma 2.

$$L(N'(i)) = \frac{(d_i - 1)L(N(i)) + 2}{d_i + 1}.$$

Proof. By definition

$$L(N'(i)) = \frac{1}{(d_i+1)d_i} \sum_{s,t \in V(N'(i)),\ s \neq t} \operatorname{dist}(s,t) = \frac{d_i - 1}{d_i + 1} L(N(i)) + \frac{2}{d_i + 1}.$$

Let's proof the relation between shortest path length in a induced subgraph and shortest path length in ambient graph if induced subgraph is obtained from ambient graph by deleting one vertex.

Theorem 1. *Let a graph G is obtained from a connected simple graph G' by deleting one vertex and $|V(G)| = n$. Then*

$$L(G') \geqslant \frac{n}{n+1} L(G),$$

where the average shortest path length $L(G)$ is defined in the ambient graph G', if G is not connected.

Proof. Let's denote the deleted vertex by v. By the triangle inequality $\forall s,t \in V(G) : \operatorname{dist}(s,v) + \operatorname{dist}(v,t) \geqslant \operatorname{dist}(s,t)$, here the equality holds if, there are no paths from s to t in G. Therefore,

$$\sum_{s,t \in V(G), s \neq t} \big(\operatorname{dist}(s,v) + \operatorname{dist}(v,t)\big) \geqslant \sum_{s,t \in V(G), s \neq t} \operatorname{dist}(s,t),$$

$$\frac{2(n-1)}{n(n-1)} \sum_{t \in V(G)} \operatorname{dist}(v,t) \geqslant \frac{1}{n(n-1)} \sum_{s,t \in V(G), s \neq t} \operatorname{dist}(s,t),$$

$$\frac{2}{n} \sum_{t \in V(G)} \operatorname{dist}(v,t) \geqslant L(G).$$

Then,

$$L(G') = \frac{1}{(n+1)n} \sum_{s,t \in V(G'), s \neq t} \mathrm{dist}(s,t)$$

$$= \frac{1}{(n+1)n} \left(2 \sum_{t \in V(G)} \mathrm{dist}(v,t) + \sum_{s,t \in V(G), s \neq t} \mathrm{dist}(s,t) \right)$$

$$= \frac{1}{n+1} \frac{2}{n} \sum_{t \in V(G)} \mathrm{dist}(v,t) + \frac{n-1}{n+1} L(G) \geqslant \frac{n}{n+1} L(G).$$

Note that if G consists of n isolated vertices then the equality holds.

We obtain corollaries.

Corollary 2. *Let a graph G is obtained from a connected simple graph $G' \subset H$ by deleting one vertex, $|V(G)| = n$ and H is connected and simple. Then*

$$L(G') \geqslant \frac{n}{n+1} L(G),$$

where the average shortest path lengths $L(G)$ and $L(G')$ are defined in the ambient graph H, if corresponded graphs are not connected.

Proof. The proof is the same as for the previous theorem and also if G consists of n isolated vertices then the equality holds.

Corollary 3.

$$L(N'(i)) \geqslant \frac{d_i}{d_i + 1} L(N(i)).$$

Let's proof the relation between shortest path length in a induced subgraph and shortest path length in ambient graph.

Theorem 2. *Let's $G' \subset G$ be induced subgraph and $|V(G)| = n, |V(G')| = n + k$. Then*

$$L(G') \geqslant \frac{n}{n+k} L(G),$$

where the average shortest path length $L(G)$ is defined in the ambient graph G', if G is not connected.

Proof. Let's construct the graph G' from G by adding sequentially k vertices and corresponded edges. Let's first sequentially add vertices adjacent to vertices of the graph G. Adding these vertices one by one we obtain a sequence of graphs $G_1, G_2, \ldots, G_p$, where $|V(G_i)| = n + i$. Further, let's add is the same way vertices adjacent to vertices of the graph G_p and so on. In the end we obtain the graph G'. By previous corollary

$$L(G') \geqslant \frac{n+k-1}{n+k} L(G_{k-1}) \geqslant \frac{n+k-1}{n+k} \frac{n-k-2}{n-k-1} L(G_{k-2})$$

$$= \frac{n-k-2}{n+k} L(G_{k-2}) \geqslant \cdots \geqslant \frac{n}{n+k} L(G).$$

Let's proof the relation between average closeness centrality and average shortest path length.

Lemma 3.

$$L(G) \geqslant \frac{n}{\sum_{v \in V(G)} \mathrm{Clo}(v)}.$$

Proof. By the inequality of harmonic mean and arithmetic mean

$$\frac{1}{n}\sum_{v\in V(G)}\mathrm{Clo}(v)=\frac{1}{n}\sum_{v\in V(G)}\frac{n-1}{\sum\limits_{t\in V(G)}\mathrm{dist}(v,t)}\geqslant\frac{n(n-1)}{\sum\limits_{v,t\in V(G)}\mathrm{dist}(v,t)}=\frac{1}{L(G)}.$$

Note that an equality holds when all average shortest path lengths from any vertex to all remaining vertices are equal.

Let's proof the relation between average shortest path length and average radiality.

Lemma 4.

$$L(G)=\mathrm{diam}(G)+1-\frac{1}{n}\sum_{v\in V(G)}\mathrm{Rad}(v).$$

Proof. The proof holds from definition

$$\frac{1}{n}\sum_{v\in V(G)}\mathrm{Rad}(v)=\frac{1}{n}\sum_{v\in V(G)}\frac{(n-1)(\mathrm{diam}(G)+1)-\sum\limits_{t\in V(G),\,t\neq v}\mathrm{dist}(v,t)}{n-1}$$

$$=\mathrm{diam}(G)+1-L(G).$$

Let's prove a theorem about a connection between the average stress centrality and average shortest path length for geodetic graphs.

Theorem 3. *If G is geodetic, then*

$$L(G)=1+\frac{1}{n(n-1)}\sum_{i\in V(G)}\mathrm{Str}(i).$$

Proof. Let's define

$$\chi_{st}(i)=\begin{cases}1, & \text{if } i\neq s\neq t - \text{is the vertex of the shortest path between } s \text{ and } t,\\ 0, & \text{otherwise.}\end{cases}$$

Thus, $\mathrm{Str}(i)=\sum\limits_{s,t\in V(G)}\chi_{st}(i)$. If there exists the unique shortest path between any two vertices in G, then $\mathrm{dist}(s,t)=\sum\limits_{i\in V(G)}\chi_{st}(i)+1$ (otherwise, $\mathrm{dist}(s,t)\leqslant\sum\limits_{i\in V(G)}\chi_{st}(i)+1$). Therefore, for any i

$$\sum_{s,t\in V(G)}\mathrm{dist}(s,t)=2|E|+\sum_{s,t\in V(G),\,\mathrm{dist}(s,t)\geqslant 2}\mathrm{dist}(s,t)$$

$$=2|E|+\sum_{s,t\in V(G),\,\mathrm{dist}(s,t)\geqslant 2}\left(\sum_{i\in V(G)}\chi_{st}(i)+1\right)$$

$$=2|E|+\sum_{i\in V(G)}\mathrm{Str}(i)+n(n-1)-2|E|=\sum_{i\in V(G)}\mathrm{Str}(i)+n(n-1).$$

Corollary 4. *For any simple connected G*

$$L(G)\leqslant 1+\frac{1}{n(n-1)}\sum_{i\in V(G)}\mathrm{Str}(i).$$

First let's proof theorem about a connection between average clustering coefficient and stress centrality and we will use it for a connection between the average shortest path length and the average local clustering coefficient further.

Theorem 4. *If there is no pendant vertices in graph, then*

$$C_{WS}(G) \geqslant \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\mathrm{Str}(i)}{d_i(d_i-1)}\right).$$

Proof. Note that $\forall j, k \in N(i) : (j,k) \notin E(N(i))$ the shortest path between j and k is $j \to i \to k$. Therefore,

$$\mathrm{Str}(i) \geqslant 2\left(\frac{d_i(d_i-1)}{2} - |E(N(i))|\right),$$

$$\frac{1}{d_i(d_i-1)}\mathrm{Str}(i) \geqslant 1 - c_i.$$

Averaging by i

$$C_{WS}(G) \geqslant \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\mathrm{Str}(i)}{d_i(d_i-1)}\right).$$

Note that for $diam(G) = 2$ holds an equality.

Corollary 5. *For any simple connected graph G*

$$C_{WS}(G) \geqslant \frac{1}{n} \sum_{i \in V(G)} \left(1 - \frac{\mathrm{Str}(i)}{d_i(d_i-1)}\right) - \frac{\textit{number of pendant vertices}}{n}.$$

Proof. Let's define $\frac{Str(i)}{d_i(d_i-1)}$ as 0 for every pendant vertex i, then in the inequality $\frac{1}{d_i(d_i-1)}Str(i) \geqslant 1 - c_i$ in the right side will be 1 and in the left side 0, thus if we add in the left side 1 for every pendant vertex, will be right equality.

Now let's prove a theorem about a connection between the average shortest path length and the average local clustering coefficient for geodetic graphs.

Theorem 5. *If G is geodetic and* $\forall i, j \in V(G)$ *hold, if* $d_i \leqslant d_j$ *then* $\mathrm{Str}(i) \leqslant \mathrm{Str}(j)$, *then*

$$1 - C_{WS}(G) \leqslant \frac{1}{n} \sum_{i \in V(G)} \frac{(L(G)-1)(n-1)}{d_i(d_i-1)} + \frac{\textit{number of pendant vertices}}{n}.$$

Proof. Let's re-numerate vertices such that $\forall i \leqslant j : d_i \leqslant d_j$. Then for $i \leqslant j$ hold $\mathrm{Str}(i) \leqslant \mathrm{Str}(j)$ and $d_i(d_i-1) \leqslant d_j(d_j-1)$. By Theorem 3 and Corollary 5 for the case if there is no pendant vertices in graph holds

$$1 - C_{WS}(G) \leqslant \frac{1}{n} \sum_{i \in V(G)} \frac{\mathrm{Str}(i)}{d_i(d_i-1)}$$

$$\overset{\text{Chebyshev's sum inequality}}{\leqslant} \left(\frac{1}{n} \sum_{i \in V(G)} \mathrm{Str}(i)\right)\left(\frac{1}{n} \sum_{i \in V(G)} \frac{1}{d_i(d_i-1)}\right)$$

$$= \frac{1}{n} \sum_{i \in V(G)} \frac{(L(G)-1)(n-1)}{d_i(d_i-1)}.$$

For pendant vertices we should add $\frac{\text{number of pendant vertices}}{n}$ in the right side. Note that if there exist two vertices $i, j \in V(G)$ such that $d_i < d_j$ and $\mathrm{Str}(i) < \mathrm{Str}(j)$ then the inequality in this theorem will be strict.

Example 1. Let's consider a star G with $V(G) = n+1$ vertices. The star is geodetic graph. The central vertex has degree n, local clustering coefficient $c_i = 0$ and stress centrality $\mathrm{Str}(i) = n(n-1)$. Other vertices are pendant ($d_i = 1,\ c_i = 0,\ \mathrm{Str}(i) = 0$). Thus for this graph holds Theorem 5.

$$L(G) = \frac{n(2n-1)+n}{(n+1)n} = \frac{2n}{n+1}, \quad C_{WS} = 0.$$

$$1 - C_{WS}(G) = 1 = \frac{\frac{n-1}{n+1}n}{n(n-1)} + \frac{n}{n+1}$$

$$= \frac{1}{n+1}\sum_{i\in V(G)} \frac{(L(G)-1)n}{d_i(d_i-1)} + \frac{\text{number of pendant vertices}}{n+1}.$$

Thus, for this example holds equality.

REFERENCES

1. Travers, J. and Milgram, S., An experimental study of the small world problem, *Social networks. Academic Press*, 1977, pp. 179–197.

2. Watts, D.J. and Strogatz, S.H., Collective dynamics of 'small-world'networks, *Nature*, 1998, vol. 393, no. 6684, pp. 440–442.

3. Borgatti, S.P. and Everett, M.G., A graph-theoretic perspective on centrality, *Social networks*, 2006, vol. 28, no. 4, pp. 466–484.

4. Kiss, C. and Bichler, M., Identification of influencers-measuring influence in customer networks, *Decision Support Systems*, 2008, vol. 46, no. 1, pp. 233–253.

5. Lee, S.H.M., Cotte, J., and Noseworthy, T.J., The role of network centrality in the flow of consumer influence *J. Consum. Psychol*, 2010, vol. 20, no. 1, pp. 66–77.

6. You, J., Gomes-Selman, J.M., Ying, R., et al., Identity-aware graph neural networks, *Proceedings of the AAAI conference on artificial intelligence*, 2021, vol. 35, no. 12, pp. 10737–10745.

7. Yuan, M.M., Guo, X., Wu, L., et al., Climate warming enhances microbial network complexity and stability, *Nature Climate Change*, 2021, vol. 11, no. 4, pp. 343–348.

This paper was recommended for publication by A.G. Kushner, a member of the Editorial Board