Journal of Reliability and Statistical Studies (ISSN: 0974-8024)
Vol. 3, Issue 1(2010): 13-29

EDGE ESTIMATION IN POPULATION OF PLANER GRAPHS USING SAMPLING

D. Shukla[1], Yashwant Singh Rajput and Narendra Singh Thakur[2]
Deptt. of Mathematics and Statistics, University of Sagar, Sagar, M.P., India.
E Mail: 1. diwakarshukla@rediffmail.com, 2. nst_stats@yahoo.co.in

Abstract

Consider a population which contains graphical relationship between two variables. There are two graphs of vertices and edges, each edge contains a length value and linked with two vertices (nodes). Mean length of all edges is unknown which is a problem to explore. This paper takes into account two planer graphs in particular, one of them is under main interest and other is an auxiliary graph. A sample of some nodes is drawn by simple random sampling (SRSWOR) along with a laid down node-sampling procedure and a class of estimators is proposed to estimate the mean length of an edge of planer graph using the structure of other planer graph as an auxiliary source of information. Optimal properties of estimators are derived and results are numerically supported with the calculation of length estimates and confidence intervals.

Keywords : Graph, Planer Graph, Edge, Vertices(nodes), Simple Random Sampling without replacement (SRSWOR), Class, Estimator, Bias, Mean Squared Error (MSE), Optimum Choice, Confidence intervals.

1. Introduction

Consider a graph $G_2 = (V, E, \psi)$ where set V consists of the five vertices V_1, V_2, V_3, V_4, and V_5; set E has eight edges (none of which is in V) like

$$V = \{V_1, V_2, V_3, V_4, V_5\} \;; \quad E = \{e_{12}, e_{13}, e_{14}, e_{23}\, e_{24}, e_{25}, e_{35}, e_{45}\}$$

The ψ is a set containing the relationship between V and E in the form of mapping. Likewise, define another graph $G_1 = (V', E', \psi')$ where V' and E' consist of

$$V' = \{V_1', V_2', V_3', V_4', V_5'\} \quad E' = \{e_{12}', e_{14}', e_{23}', e_{35}', e_{45}'\}$$

and ψ' has relationship of V' and E',

Note 1.1: The graph G_2 (or G_1) is said to be a planer graph if there exists some geometric representation of G_2 (or G_1) which can be drawn on a plane such that no two of edges can intersect to each other [see Deo(2001)]. Detail description of planer graph is in Parson (1971). Some useful research contributions to planer graphs are due to Frederickson (1988), Gazot and Reif (1990), Shih et al. (1990), Grigni et al. (1995), Osthus et al. (2003), Aleksandrav et al.(2007), etc.

Fig 1 is showing a structure of two linked planer graphs G_1 and G_2 (with a common vertex $V_2 = V_2'$). Suppose G_1 and G_2 together constitute a population of vertices (nodes) (V, V') and edges (E, E'). Overall average length of edges in G_1 is assumed known but overall mean edge length is not known for G_2. We want to estimate this using mean edge information of graph G_1 and with the help of a random sample of some nodes drawn according to following node-sampling procedure.

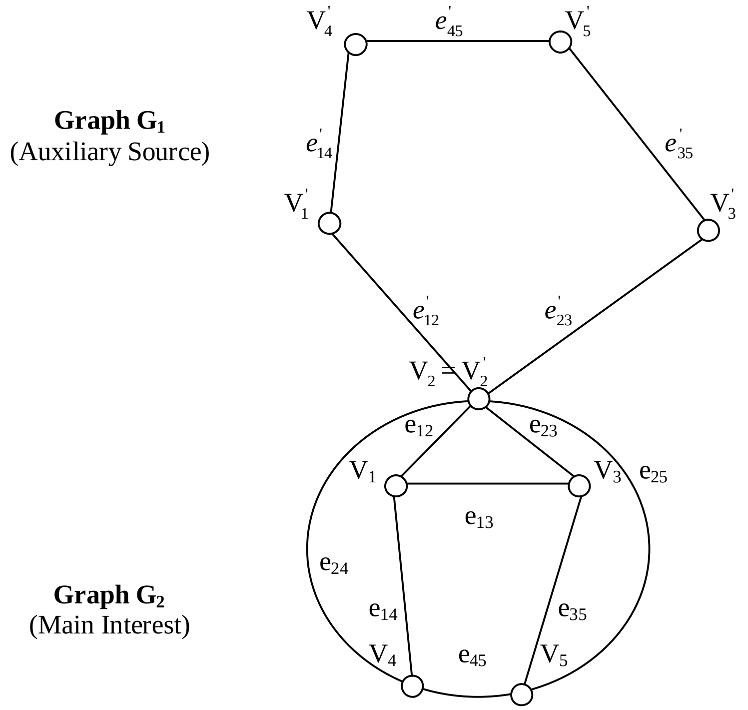

Fig. 1: A structure of two linked planer graphs G_1 and G_2

1.1 Node- Sampling Procedure

Suppose G_1 and G_2 both individually have equal nodes M in form of V_i and $V_i^{'}$ (i=1,2,3,………M).

Step I: Construct node-edge (NE) table for the population of M units. Common vertex represents both in G_1 and G_2 with the respective group edges (like V_2 in $G_2 = V_2^{'}$ in G_1).

Step II : Construct false-node-edge (FNE) matrix for G_1 and G_2 separately assuming that (i, j)th element of matrix is unity if the jth edge is associated (or linked) with ith vertex (node), otherwise zero. Some false edges are added in FNE matrix in order to maintain equal number of edges in both G_1 and G_2, by the diagonal-repetition of edges of same group on priority basis. If main diagonal is full, the very next upper diagonal is taken into consideration for the false edge creation until equality of edges in G_1 and G_2 is met out. At the end, total number of edges in G_1 and G_2 are same, equal to N, shown below as an example for Fig. 1 **:**

Graph G_2		**Graph G_1**	
Vertex	**Edge**	**Vertex**	**Edge**
V_1	e_{12}, e_{13}, e_{14}	V_1'	e_{12}', e_{14}'
V_2	e_{12}, e_{23}, e_{24}, e_{25}	V_2'	e_{12}', e_{23}'
V_3	e_{13}, e_{23}, e_{35}	V_3'	e_{23}', e_{35}'
V_4	e_{14}, e_{24}, e_{45}	V_4'	e_{14}', e_{45}'
V_5	e_{25}, e_{35}, e_{45}	V_5'	e_{35}', e_{45}'

Table 1: Node-Edge Table (NE Table) For Fig. 1

Nodes	**Edges**								**False edges**								*Row count*	*mean edge length*
	e_{12}	e_{13}	e_{14}	e_{23}	e_{24}	e_{25}	e_{35}	e_{45}	e_{12}	e_{13}	e_{14}	e_{23}	e_{24}	e_{25}	e_{35}	e_{45}		
V_1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	3	$\bar{e}_1$
V_2	1	0	0	1	1	1	0	0	0	0	0	0	0	0	0	0	4	$\bar{e}_2$
V_3	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	0	3	$\bar{e}_3$
V_4	0	0	1	0	1	0	0	1	0	0	0	0	0	0	0	0	3	$\bar{e}_4$
V_5	0	0	0	0	0	1	1	1	0	0	0	0	0	0	0	0	3	$\bar{e}_5$
																Total	**16**	

Table 2: FNE Matrix for Graph G_2 (FNEM)

Nodes	**Edges**					**False Edges**					*Row count*	*mean edge length*
	e_{12}'	e_{14}'	e_{23}'	e_{35}'	e_{45}'	e_{12}'	e_{14}	e_{23}	e_{35}	e_{45}		
V_1'	1	1	0	0	0	1	1	0	0	0	4	$\bar{e}_1'$
V_2'	1	0	1	0	0	0	1	0	0	0	3	$\bar{e}_2'$
V_3'	0	0	1	1	0	0	0	1	0	0	3	$\bar{e}_3'$
V_4'	0	1	0	0	1	0	0	0	1	0	3	$\bar{e}_4'$
V_5'	0	0	0	1	1	0	0	0	0	1	3	$\bar{e}_5'$
										Total	**16**	

Table 3: FNE Matrix for Graph G_1 (FNEM)

For the i^{th} node of FNEM, row-wise mean-length edge is denoted by $\overline{e_i}$ and $\overline{e_i^{'}}$ based on their counts n_i and $n_i^{'}$ respectively for graph G_1 and G_2. The overall population means are**:**

$$\left.\begin{aligned} \overline{e}^{(2)} &= \frac{1}{N}\sum_{i=1}^{N} n_i \overline{e_i} \quad \text{for graph } G_2 \\ \overline{e}^{(1)} &= \frac{1}{N}\sum_{i=1}^{N} n_i^{'} \overline{e_i^{'}} \quad \text{for graph } G_1 \end{aligned}\right\} \tag{1.1}$$

where $N = \sum_i n_i^{'} = \sum_i n_i$ in FNEM

Step III : Draw a sample of n nodes (n < M) from graph G_2 by SRSWOR and choose the corresponding similar number of nodes in G_1. For example, if V_3 from G_2 appears then $V_3^{'}$ of G_1 appears automatically, in the sample.

Step IV : For any i^{th} sampled node (vertex), pickup i^{th} row of FNEM of G_2 and G_1 both. Select very first edge from left among all non-zero elements of the column, say j, of i^{th} row separately for both G_1 and G_2. This provides two edge-lengths e_{ij} and $e_{ij}^{'}$ related to the i^{th} node for sample.

Step V : Continue the procedure laid down in step IV, for all nodes i (i=1,2,3….n), appeared in the sample using SRSWOR.

The end of Node Sampling procedure provides a random sample of n edge lengths, drawn from N edges, in the form of e_{si} and $e_{si}^{'}$ related to i^{th} node.

The sample mean of edge lengths are **:**

$$\left.\begin{aligned} \overline{e}_s^{(2)} &= \frac{1}{n}\sum_{i=1}^{n} e_{si} \quad \text{for graph } G_2 \\ \overline{e}_s^{(1)} &= \frac{1}{n}\sum_{i=1}^{n} e_{si}^{'} \quad \text{for graph } G_1 \end{aligned}\right\} \tag{1.2}$$

2. A Class of Edge Estimators

Deriving a motivation from Singh and Shukla (1987, 1993), Shukla et al. (1991) and Shukla (2002), assume the mean length $\overline{e}^{(1)}$ for G_1 is known ; $\overline{e}^{(2)}$ for G_2 is unknown and under target of estimation. The discussion over variable of main interest and use of auxiliary information, for estimation purpose, is in Sukhatme et al. (1984), Singh & Choudhary (1986), Cochran (2005) etc.

Define $\bar{e}_s^{(*)} = \left[\dfrac{N\bar{e}^{(1)} - n\bar{e}_s^{(1)}}{N-n}\right]$; $f = n/N$; (2.1)

$$\bar{e}_s^{(a)} = \left[(1-f)\bar{e}^{(1)} - f\,\bar{e}_s^{(1)}\right] \tag{2.2}$$

$$\bar{e}_s^{(b)} = \left[f(\bar{e}_s^{(*)}) + (1-f)\bar{e}_s^{(a)}\right] \tag{2.3}$$

Using equations (2.1), (2.2), (2.3), the proposed class of edge-estimators M_e is

$$M_e = \bar{e}_s^{(2)}\left[\frac{A\bar{e}^{(1)} + C\bar{e}_s^{(b)} + f\,B\bar{e}_s^{(*)} + C\bar{e}_s^{(a)}}{A\bar{e}^{(1)} + C\bar{e}_s^{(b)} + C\bar{e}_s^{(*)} + f\,B\bar{e}_s^{(a)}}\right] \tag{2.4}$$

where A = (k-1)(k -2); B = (k -1)(k -4); C = (k -2)(k -3)(k -4)

The term k is a suitably chosen positive constant ($0 < k < \infty$).

Note 2.1: The class (2.4) is close to the structure of factor-type estimator discussed by Singh and Shukla (1987), Shukla (2002). Different graphical structures which a population may contain, are described in Dev (2001).

Note 2.2: The (2.4) contains an unknown parameter k, whose different choices generate a series of mean-edge estimators. Therefore, it could be looked upon as a class of edge-estimators to estimate unknown $\bar{e}^{(2)}$.

2.1 Special Estimators

At $k = 1$ $\quad [M_e]_1 = \bar{e}_s^{(2)}\left[\dfrac{\bar{e}_s^{(b)} + \bar{e}_s^{(a)}}{\bar{e}_s^{(b)} + \bar{e}_s^{(*)}}\right]$ (2.5)

$k = 2$ $\quad [M_e]_2 = \bar{e}_s^{(2)}\left[\dfrac{\bar{e}_s^{(*)}}{\bar{e}_s^{(a)}}\right]$ (2.6)

$k = 3$ $\quad [M_e]_3 = \bar{e}_s^{(2)}\left[\dfrac{\bar{e}^{(1)} - f\bar{e}_s^{(*)}}{\bar{e}^{(1)} - f\bar{e}_s^{(a)}}\right]$ (2.7)

$k = 4$ $\quad [M_e]_4 = \bar{e}_s^{(2)}$ (2.8)

3. Setting Approximations

Suppose large number of nodes and edges are linked in a population of two planer graphs G_1 and G_2, a large sample of n vertices is drawn by node-sampling procedure described in section 1.1. For any two small positive numbers r_1 and r_2, ($|r_1| < 1$, $|r_2| < 1$), the approximation is [see Cochran (2005)]

$$\left.\begin{aligned} &\text{for graph } G_2 \quad \bar{e}_s^{(2)} = \bar{e}^{(2)}(1+r_2) \\ &\text{for graph } G_1 \quad \bar{e}_s^{(1)} = \bar{e}^{(1)}(1+r_1) \end{aligned}\right\} \quad (3.1)$$

with conditions

(i) $E(r_1) = E(r_2) = 0$ (3.2)

(ii) $E(r_2^2) = \left[\bar{e}^{(2)}\right]^{-2}\left\{E\left(\bar{e}_s^{(2)} - \bar{e}^{(2)}\right)^2\right\}$

$$= \left[\bar{e}^{(2)}\right]^{-2}\left\{V\left(\bar{e}_s^{(2)}\right)\right\}$$

$$= (N-n)(Nn)^{-1}(C_e^{(2)})^2 \quad (3.3)$$

(iii) $E(r_1^2) = \left[\bar{e}^{(1)}\right]^{-2}\left\{E\left(\bar{e}_s^{(1)} - \bar{e}^{(1)}\right)^2\right\}$

$$= \left[\bar{e}^{(1)}\right]^{-2}\left\{V\left(\bar{e}_s^{(1)}\right)\right\}$$

$$= (N-n)(Nn)^{-1}(C_e^{(1)})^2 \quad (3.4)$$

(iv) $E(r_1 r_2) = \left(\bar{e}^{(1)}.\bar{e}^{(2)}\right)^{-1}\left\{E\left(\bar{e}_s^{(1)} - \bar{e}^{(1)}\right)\left(\bar{e}_s^{(2)} - \bar{e}^{(2)}\right)\right\}$

$$= (N-n)(Nn)^{-1}(\rho.C_e^{(1)}.C_e^{(2)}) \quad (3.5)$$

Note 3.1: Symbols $\left\{S_e^{(1)}\right\}^2, \left\{S_e^{(2)}\right\}^2$ are population mean squares of N edges of graphs G_1 and G_2 as described in FNE matrices. The $V(\cdot)$ denotes variance and $E(\cdot)$ denotes expectation of the estimates based on sample n. Moreover, $\left\{C_e^{(1)}\right\}^2 = \left\{S_e^{(1)}\right\}^2\left\{\bar{e}^{(1)}\right\}^{-2}$; $\left\{C_e^{(2)}\right\}^2 = \left\{S_e^{(2)}\right\}^2\left\{\bar{e}^{(2)}\right\}^{-2}$; are coefficients of variation of N edges in both the FNEM .Further, ρ denotes correlation coefficient between N edges in FNEM.

Remark 3.1: The (2.1), (2.2) and (2.3) could be expressed in the approximate form using (3.1) as

$$\bar{e}_s^{(*)} = \bar{e}^{(1)}\left[1-\delta_1 r_1\right] \;;\; \delta_1 = n(N-n)^{-1} \quad \bar{e}_s^{(a)} = \bar{e}^{(1)}\left[1+ f r_1\right];$$

$$\bar{e}_s^{(b)} = \bar{e}^{(1)}\left[1+\delta_2 r_1\right] \;;\; \delta_2 = (1- f -\delta_1)\, f$$

Theorem 3.1: Using (3.1) and remark 3.1 the class of estimators (2.4) could be expressed in the approximate form

$$M_e = \bar{e}^{(2)}\left[1-\left(\beta^*-\alpha^*\right)r_1+\beta^*\left(\beta^*-\alpha^*\right)r_1^2+r_2-\left(\beta^*-\alpha^*\right)r_1r_2+\beta^*\left(\beta^*-\alpha^*\right)r_1^2r_2\right]$$

where $\alpha^* = \left(C\delta_2 - fB\delta_1 + Cf\right)/\Delta$; $\beta^* = \left(C\delta_2 - C\delta_1 + f^2B\right)/\Delta$;

$\Delta = \left(A + 2C + fB\right)$

Proof : Using (3.1) and remark 3.1, one can express class M_e of (2.4)

$$M_e = \bar{e}^{(2)}\left(1+r_2\right)\left[\left[\frac{\left\{A\bar{e}^{(1)} + C\bar{e}^{(1)}\left(1+\delta_2 r_1\right) + fB\bar{e}^{(1)}\left(1-\delta_1 r_1\right) + C\bar{e}^{(1)}\left(1+f\ r_1\right)\right\}}{\left\{A\bar{e}^{(1)} + C\bar{e}^{(1)}\left(1+\delta_2 r_1\right) + C\bar{e}^{(1)}\left(1-\delta_1 r_1\right) + fB\bar{e}^{(1)}\left(1+f\ r_1\right)\right\}}\right]\right]$$

$$= \bar{e}^{(2)}\left(1+r_2\right)\left[\left[\frac{\left\{\Delta + \left(C\delta_2 - fB\delta_1 + Cf\right)r_1\right\}}{\left\{\Delta + \left(C\delta_2 - C\delta_1 + f^2B\right)r_1\right\}}\right]\right]$$

$$= \bar{e}^{(2)}\left(1+r_2\right)\left(1+\alpha^* r_1\right)\left(1+\beta^* r_1\right)^{-1}$$

$$= \bar{e}^{(2)}\left(1+r_2\right)\left(1+\alpha^* r_1\right)\left(1-\beta^* r_1 + (\beta^* r_1)^2 - (\beta^* r_1)^3 + \cdots\cdots\right)$$

Assume the term $\left(\beta^* r_1\right)^j$ very small when $j > 2$, therefore, ignore all terms in expansion $\left(1+\beta^* r_1\right)^{-1}$ for $j > 2$, we get[see Sukhatme et al. (1984), Singh and Choudhary (1986), Cochran (2005)]

$$M_e = \bar{e}^{(2)}\left(1+r_2\right)\left(1+\alpha^* r_1\right)\left(1-\beta^* r_1 + (\beta^* r_1)^2\right)$$

$$= \bar{e}^{(2)}\left[\begin{array}{l}1-\left(\beta^*-\alpha^*\right)r_1+\beta^*\left(\beta^*-\alpha^*\right)r_1^2+r_2-\left(\beta^*-\alpha^*\right)r_1r_2 \\ \qquad\qquad +\beta^*\left(\beta^*-\alpha^*\right)r_1^2r_2\end{array}\right]$$

Hence the theorem .

Theorem 3.2 : Bias of the proposed class of estimators M_e is

$$B\left(M_e\right) = \bar{e}^{(2)}\left[\frac{N-n}{Nn}\right]\left[\left(\beta^*-\alpha^*\right)\left(\beta^*\left(C_e^{(1)}\right)^2 - \rho C_e^{(1)}C_e^{(2)}\right)\right]$$

Proof : Let $B(.)$ denotes the bias, then using theorem 3.1 **:**

$$E\left(M_e\right) = E\left[\bar{e}^{(2)}\left\{\begin{array}{l}1-\left(\beta^*-\alpha^*\right)r_1+\beta^*\left(\beta^*-\alpha^*\right)r_1^2+r_2-\left(\beta^*-\alpha^*\right)r_1r_2 \\ \qquad\qquad +\beta^*\left(\beta^*-\alpha^*\right)r_1^2r_2\end{array}\right\}\right]$$

$$= \bar{e}^{(2)}\left[\begin{array}{l}1-\left(\beta^*-\alpha^*\right)E\left(r_1\right)+\beta^*\left(\beta^*-\alpha^*\right)E\left(r_1^2\right)+E\left(r_2\right)-\left(\beta^*-\alpha^*\right)E\left(r_1r_2\right) \\ \qquad\qquad +\beta^*\left(\beta^*-\alpha^*\right)E\left(r_1^2r_2\right)\end{array}\right]$$

Using results in (3.2) to (3.5) and substituting

$$E\left(r_1^i . r_2^j\right) = 0 \quad \text{when } i + j \geq 3;\ i, j = 0,1,2,3 \ldots \text{[see Cochran (2005)]} \qquad (3.6)$$

we get

$$E(M_e) = \bar{e}^{(2)}\left[1+\beta^*(\beta^*-\alpha^*)\left(\frac{N-n}{Nn}\right)\left(C_e^{(1)}\right)^2-(\beta^*-\alpha^*)\left(\frac{N-n}{Nn}\right)\rho C_e^{(1)}C_e^{(2)}\right]$$

$$= \bar{e}^{(2)} + \bar{e}^{(2)}\left(\frac{N-n}{Nn}\right)\left[\beta^*(\beta^*-\alpha^*)\left(C_e^{(1)}\right)^2-(\beta^*-\alpha^*)\rho C_e^{(1)}C_e^{(2)}\right]$$

Therefore, the bias is

$$B(M_e) = \left[E(M_e)-\bar{e}^{(2)}\right]$$

$$= \bar{e}^{(2)}\left[\frac{N-n}{Nn}\right]\left[(\beta^*-\alpha^*)\left(\beta^*\left(C_e^{(1)}\right)^2-\rho C_e^{(1)}C_e^{(2)}\right)\right]$$

Hence the theorem.

Remark 3.2 : The class M_e contains a sub-class of unbiased estimators if

$$\beta^* = \rho\frac{C_e^{(2)}}{C_e^{(1)}} = V \quad (\text{Let})$$

Proof : Substituting $B(M_e)=0 \Rightarrow \beta^*\left(C_e^{(1)}\right)^2-\rho C_e^{(1)}C_e^{(2)} = 0$ and hence the result.

Remark 3.3 : The remark 3.2 generates an equation in terms of A, B, C, f and V

$$VA+f(V-f)B+(\delta_1-\delta_2+2V)C=0 \tag{3.7}$$

which provides a necessary condition for obtaining unbiased estimators in the class, up to the first order of approximation, by suitable choices of k. The (3.7) is a cubic equation in k which gives at most three values of k for which the bias is zero. One can chose the useful value of k related to lowest m.s.e.

Theorem 3.3 : Mean squared error of estimator M_e is

$$MSE(M_e)=M(M_e)=\left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2+\left(P^*\right)^2\left(C_e^{(1)}\right)^2+2P^*\rho\, C_e^{(1)}C_e^{(2)}\right]$$

where $P^* = (\alpha^*-\beta^*)$

Proof : Define $MSE(M_e)=M(M_e)=E\left(M_e-\bar{e}^{(2)}\right)^2$. On replacing M_e using theorem 3.1 together with equation (3.6), we get

$$M(M_e)=\left(\bar{e}^{(2)}\right)^2\left[E(r_2^2)+(\beta^*-\alpha^*)^2E(r_1^2)-2(\beta^*-\alpha^*)E(r_1r_2)\right]$$

Using equation (3.2) to (3.5) ,

$$M(M_e)=\left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2+(\beta^*-\alpha^*)^2\left(C_e^{(1)}\right)^2+2\rho(\beta^*-\alpha^*)C_e^{(1)}C_e^{(2)}\right]$$

$$=\left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2+\left(P^*\right)^2\left(C_e^{(1)}\right)^2+2P^*\rho\, C_e^{(1)}C_e^{(2)}\right]$$

Hence the theorem.

Remark 3.4 : Some special cases related to bias and m.s.e. are

$$At \quad k=1, \; B(M_e)_1 = \bar{e}^{(2)}\left[\frac{N-n}{Nn}\right]\left[\left(\frac{\delta_1+f}{2}\right)\left(\left(\frac{\delta_1-\delta_2}{2}\right)\left(C_e^{(1)}\right)^2 + \rho\, C_e^{(1)}C_e^{(2)}\right)\right]$$

$$M(M_e)_1 = \left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2 + \left(\frac{\delta_1+f}{2}\right)^2\left(C_e^{(1)}\right)^2 + 2\left(\frac{\delta_1+f}{2}\right)\rho\, C_e^{(1)}C_e^{(2)}\right]$$

$$At \quad k=2, \; B(M_e)_2 = \bar{e}^{(2)}\left[\frac{N-n}{Nn}\right]\left[(f+\delta_1)\left(f\left(C_e^{(1)}\right)^2 - \rho\, C_e^{(1)}C_e^{(2)}\right)\right]$$

$$M(M_e)_2 = \left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2 + (f+\delta_1)^2\left(C_e^{(1)}\right)^2 - 2(f+\delta_1)\rho\, C_e^{(1)}C_e^{(2)}\right]$$

$$At \quad k=3, \; B(M_e)_3 = \bar{e}^{(2)}\left[\frac{N-n}{Nn}\right]\left[\left(\frac{f(f+\delta_1)}{f-1}\right)\left(\left(\frac{f^2}{f-1}\right)\left(C_e^{(1)}\right)^2 - \rho\, C_e^{(1)}C_e^{(2)}\right)\right]$$

$$M(M_e)_3 = \left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2 + \left(\frac{f(f+\delta_1)}{f-1}\right)^2\left(C_e^{(1)}\right)^2 - 2\left(\frac{f(f+\delta_1)}{f-1}\right)\rho\, C_e^{(1)}C_e^{(2)}\right]$$

$$At \quad k=4, \; B(M_e)_4 = 0$$

$$M(M_e)_4 = \left(\bar{e}^{(2)}\right)^2\left[\frac{N-n}{Nn}\right]\left[\left(C_e^{(2)}\right)^2\right]$$

4. Optimum Choices of *k*

Expression of mean square error of the class depends on parameter P^* which is a function of k. One can obtain the appropriate choice of P^* subject to condition the mean squared error is minimum.

Theorem 4.1 : The minimum mean squared error occurs when

$$P^* = -V$$

Proof : Differentiating $MSE(M_e)$ of theorem 3.3 with respect to P^* and equating to zero,

$$\frac{d}{dP^*}\left[M(M_e)\right] = 0 \quad \Rightarrow P^*\left(C_e^{(1)}\right)^2 + \rho\left(C_e^{(1)}\right)\left(C_e^{(2)}\right) = 0$$

$$P^* = -\frac{\rho\left(C_e^{(2)}\right)}{\left(C_e^{(1)}\right)} = -V \qquad (4.1)$$

Hence the theorem.

Remark 4.1 : The optimality condition (4,1) provides an equation,

$$VA + f\left[V - f - \delta_1\right]B + \left[2V + f + \delta_1\right]C = 0 \qquad (4.2)$$

which is cubic in term of parameter k and for known values of f and V, there are at most three values of k for which the m.s.e. could be optimized (minimized).

Remark 4.2 : Let $k_1^{'}$, $k_2^{'}$ and $k_3^{'}$ be three values for which $MSE(M_e)$ is minimum using equation (4.2).The best choice among them is,

$$k'_{opt} = Min\left[B(M_e)_{k'_1}, B(M_e)_{k'_2}, B(M_e)_{k'_3}\right] \tag{4.3}$$

5. Numerical Illustrations

Consider graphical population, described in Fig. 2, containing G_1 and G_2. Both are planer graphs, G_2 is of main interest, G_1 an auxiliary source. Both are related to each other by a common vertex $(V_1 = V'_1)$, therefore, it is worth to assume a correlation between them. Aim is to estimate an average edge length of G_2 using the known information of edges in G_1, with a sample drawn by node sampling procedure. The total edges in FNEM are *N*=32 for both graphs, total vertices are *M*=9 and sample size is *n*=4.

S. No.	Sample Vertices	For Graph G_1		For Graph G_2	
		Sample Edge	$\bar{e}_s^{(1)}$	Sample Edge	$\bar{e}_s^{(2)}$
1	(V_1, V_2, V_3, V_4)	$e'_{12} = 15, e'_{12} = 15,$ $e'_{13} = 14, e'_{14} = 19$	15.75	$e_{12} = 6, e_{12} = 6,$ $e_{13} = 8, e_{14} = 8$	7.00
2	(V_1, V_2, V_4, V_5)	$e'_{12} = 15, e'_{12} = 15,$ $e'_{14} = 19, e'_{15} = 12$	15.25	$e_{12} = 6, e_{12} = 6,$ $e_{14} = 8, e_{15} = 9$	7.25
3	(V_2, V_3, V_4, V_5)	$e'_{12} = 15, e'_{13} = 14,$ $e'_{14} = 19, e'_{15} = 12$	15.00	$e_{12} = 6, e_{13} = 8,$ $e_{14} = 8, e_{15} = 9$	7.75
4	(V_2, V_3, V_5, V_6)	$e'_{12} = 15, e'_{13} = 14,$ $e'_{15} = 12, e'_{46} = 14$	14.00	$e_{12} = 6, e_{13} = 8,$ $e_{15} = 9, e_{46} = 8$	7.75
5	(V_3, V_4, V_5, V_7)	$e'_{13} = 14, e'_{14} = 19,$ $e'_{15} = 12, e'_{47} = 18$	15.75	$e_{13} = 8, e_{14} = 8,$ $e_{15} = 9, e_{47} = 9$	8.50
6	(V_4, V_5, V_6, V_8)	$e'_{14} = 19, e'_{15} = 12,$ $e'_{46} = 14, e'_{58} = 17$	15.50	$e_{14} = 8, e_{15} = 9,$ $e_{46} = 8, e_{58} = 6$	7.75
7	(V_5, V_6, V_7, V_9)	$e'_{15} = 12, e'_{46} = 14,$ $e'_{47} = 18, e'_{59} = 16$	15.50	$e_{15} = 9, e_{46} = 8,$ $e_{47} = 9, e_{59} = 5$	7.75
8	(V_6, V_7, V_8, V_9)	$e'_{46} = 14, e'_{47} = 18,$ $e'_{58} = 17, e'_{59} = 16$	16.25	$e_{46} = 8, e_{47} = 9,$ $e_{58} = 6, e_{59} = 5$	7.00

Table 4: Sample Edge Description for N=4 Drawn as per Node Sampling Procedure

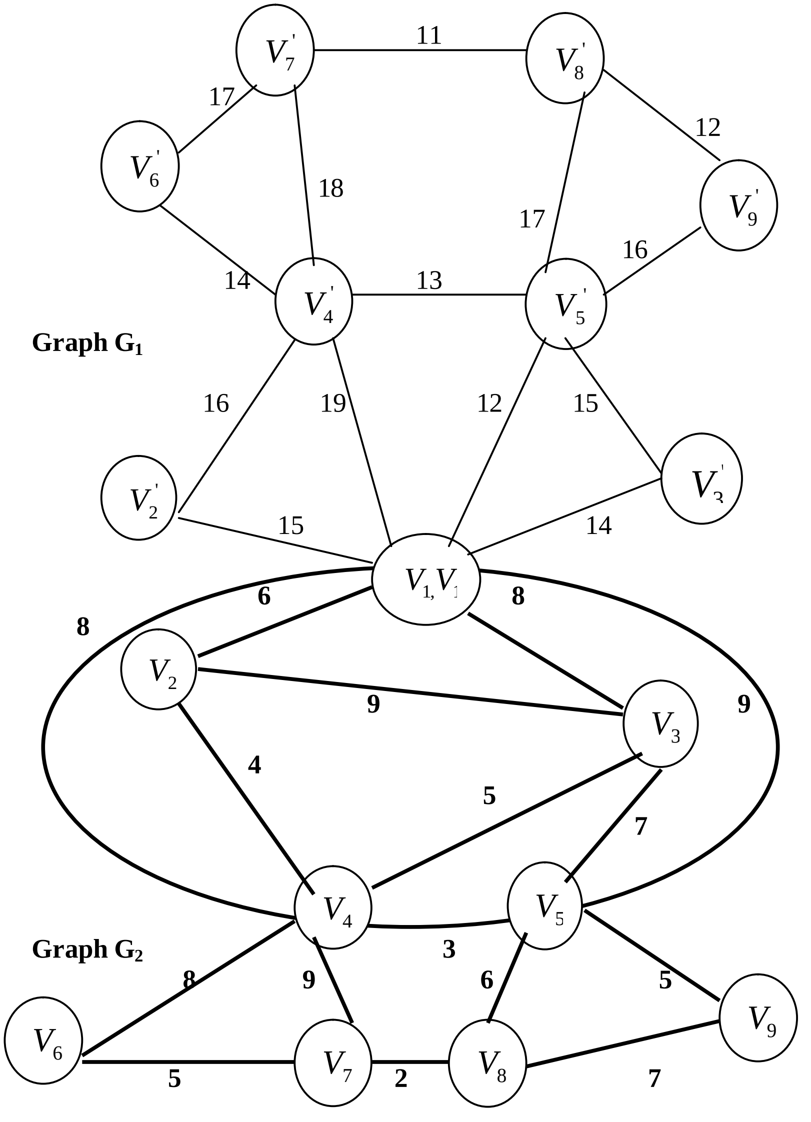

Fig. 2: Graphical population containing G_1 and G_2

The FNE matrices for G_1 and G_2, as per Fig. 2, are given in the Tables 5 and 6 respectively.

Edges

Nodes

	e'_{12}	e'_{13}	e'_{14}	e'_{15}	e'_{24}	e'_{35}	e'_{45}	e'_{46}	e'_{47}	e'_{58}	e'_{59}	e'_{67}	e'_{78}	e'_{89}
V'_1	1	1	1	1	0	0	0	0	0	0	0	0	0	0
V'_2	1	0	0	0	1	0	0	0	0	0	0	0	0	0
V'_3	0	1	0	0	0	1	0	0	0	0	0	0	0	0
V'_4	0	0	1	0	1	0	1	1	1	0	0	0	0	0
V'_5	0	0	0	1	0	1	1	0	0	1	1	0	0	0
V'_6	0	0	0	0	0	0	0	1	0	0	0	1	0	0
V'_7	0	0	0	0	0	0	0	0	1	0	0	1	1	0
V'_8	0	0	0	0	0	0	0	0	0	1	0	0	1	1
V'_9	0	0	0	0	0	0	0	0	0	0	1	0	0	1

False edges

e'_{12}	e'_{13}	e'_{14}	e'_{15}	e'_{23}	e'_{35}	e'_{45}	e'_{46}	e'_{47}	e'_{58}	e'_{59}	e'_{67}	e'_{78}	e'_{89}	*count*	*Mean* edge length
1	0	0	0	0	0	0	0	0	0	0	0	0	0	5	15.000
0	1	0	0	0	0	0	0	0	0	0	0	0	0	3	15.000
0	0	1	0	0	0	0	0	0	0	0	0	0	0	3	16.000
0	0	0	1	0	0	0	0	0	0	0	0	0	0	6	15.333
0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	14.600
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	15.500
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	15.333
0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	13.333
0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	14.000
													Total	**32**	

Table 5: FNE Matrix for G_1

Edges

Nodes	e_{12}	e_{13}	e_{14}	e_{15}	e_{23}	e_{24}	e_{34}	e_{35}	e_{45}	e_{46}	e_{47}	e_{58}	e_{59}	e_{67}	e_{78}	e_{89}
V_1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0
V_2	1	0	0	0	1	1	0	0	0	0	0	0	0	0	0	0
V_3	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	0
V_4	0	0	1	0	0	1	1	0	1	1	1	0	0	0	0	0
V_5	0	0	0	1	0	0	0	1	1	0	0	1	1	0	0	0
V_6	0	0	0	0	0	0	0	0	0	1	0	0	0	1	0	0
V_7	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1	0
V_8	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1	1
V_9	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	1

False edges

e_{12}	e_{13}	e_{14}	e_{15}	e_{23}	e_{24}	e_{34}	e_{35}	e_{45}	e_{46}	e_{47}	e_{50}	e_{59}	e_{67}	e_{78}	e_{89}	*count*	*Mean* edge length
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	7.750
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	6.333
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	4	7.250
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	6	6.167
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	5	6.000
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6.500
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5.333
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3	5.000
0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	6.000

Total 32

Table 6: FNE Matrix for G_2

GRAPH G_2						GRAPH G_1					
Node	**Edges**	**Length**	**False edge**	**Len-gth**	**Average length**	**Node**	**Edges**	**Length**	**False edge**	**Length**	**Average length**
V_1	$e_{12}, e_{13}, e_{14}, e_{15}$	6,8, 8,9	-	-	7.750	V_1'	$e_{12}', e_{13}', e_{14}', e_{15}'$	15,14, 19,12	e_{12}'	15	15.000
V_2	e_{12}, e_{23}, e_{24}	6,9, 4	-	-	6.333	V_2'	e_{12}', e_{24}'	15,16	e_{13}'	14	15.000
V_3	$e_{13}, e_{23}, e_{34}, e_{35}$	8,9, 5,7	-	-	7.250	V_3'	e_{13}', e_{35}'	14,15	e_{14}'	19	16.000
V_4	$e_{14}, e_{24}, e_{34} e_{45}, e_{46}, e_{47}$	8,4, 5,3, 8,9	-	-	6.167	V_4'	$e_{14}', e_{24}', e_{45}', e_{46}', e_{47}'$	19,16, 13,14, 18	e_{15}'	12	15.333
V_5	$e_{15}, e_{35}, e_{45}, e_{58}, e_{59}$	9,7, 3,6, 5	-	-	6.000	V_5'	$e_{15}', e_{35}', e_{45}', e_{58}', e_{59}'$	12,15, 13,17, 16	-	-	14.600
V_6	$e_{46}, e_{67},$	8,5	-	-	6.500	V_6'	e_{46}', e_{67}'	14,17	-	-	15.500
V_7	e_{47}, e_{67}, e_{78}	9,5, 2	-	-	5.333	V_7'	$e_{47}', e_{67}', e_{78}'$	18,17, 11	-	-	15.333
V_8	e_{58}, e_{78}, e_{89}	6,2, 7	-	-	5.000	V_8'	$e_{58}', e_{78}', e_{89}'$	17,11, 12	-	-	13.333
V_9	e_{59}, e_{89}	5,7	-	-	6.000	V_9'	e_{59}', e_{89}'	16,12	-	-	14.000

Table 7: Edge Vertex Description of Population

S.No.	Parameter	Value	S.No.	Parameter	Value
1	M	9	7	$\left(C_e^{(2)}\right)^2$	0.115662
2	N	4	8	$\left(C_e^{(1)}\right)^2$	0.025426
3	$\bar{e}^{(2)}$	6.3125	9	$C_e^{(2)}$	0.340092
4	$\bar{e}^{(1)}$	14.9375	10	$C_e^{(1)}$	0.159457
5	$\left(S_e^{(2)}\right)^2$	4.608871	11	ρ	0.130111
6	$\left(S_e^{(1)}\right)^2$	5.673387	12	V	0.277502

Table 8: Population Parameter Description

S.No	Value of k	$Bias[.]_{k_i}$ at $k=k_i$	MSE $[.]_{k_i}$ at $k=k_i$
1	$k_1 = 2.001256$	$B[M_e]_{k_1} = -0.000115$	$MSE[M_e]_{k_1} = 0.115192$
2	$k_2 =$ --	$B[M_e]_{k_2} =$ --	$MSE[M_e]_{k_2} =$ --
3	$k_3 =$ --	$B[M_e]_{k_3} =$ --	$MSE[M_e]_{k_3} =$ --

Table 9: k-Value When Estimator is Unbiased (With *MSE*)

Using (3.7), the unbiased estimator is achievable in the class when values of k are according to Table 9.

S.No	Value of k'	*Bias* $[.]_{k_i'}$ at *k*= k_i'	*MSE* $[.]_{k_i'}$ at *k*= k_i'
1	$k_1' = 2.026643$	$B[M_e]_{k_1'}$ = -0.000185	$MSE[M_e]_{k_1'} = 0.114661$
2	$k_2' =$ --	$B[M_e]_{k_2'} =$ --	$MSE[M_e]_{k_2'} =$ --
3	$k_3' =$ --	$B[M_e]_{k_3'} =$ --	$MSE[M_e]_{k_3'} =$ --

Table 10: k -Value for Optimum MSE (M_e)

Using (4.2), the optimum estimator could be obtained in the class when values of k are according to Table 10.

k	*Bias* (M_e)	*MSE* (M_e)	*V* (M_e)
k =1	$B[M_e]_1 = 0.0014$	$MSE[M_e]_1 = 1.0286$	$V[M_e]_1 = 1.0285$
k=2	$B[M_e]_2$ = -0.0014	$MSE[M_e]_2 = 0.9911$	$V[M_e]_2 = 0.9910$
k=3	$B[M_e]_3$ = -0.0002	$MSE[M_e]_3 = 1.0135$	$V[M_e]_3 = 1.0134$
$k = k_1$ =2.001256 (unbiased)	$B[M_e]_{k_1}$ = -0.0001	$MSE[M_e]_{k_1} = 0.1152$	$V[M_e]_{k_1} = 0.1151$
$k = k_{opt} = k_1'$ =2.026643 (opt *MSE*)	$B[M_e]_{k_1'}$ = -0.0002	$MSE[M_e]_{k_1'} = 0.1147$	$V[M_e]_{k_1'} = 0.1147$

Table 11: Calculation of *MSE* (M_e) for various values of *k*

where variance is computed by $V(.) = MSE(.) - [Bais(.)]^2$.

N=32, M=9, n=4, $\bar{e}^{(1)}$ =14.9375, $\bar{e}^{(2)}$ =6.3125 (which is actually unknown)							
S.No	**Sample Vertices**	**Sample Means**		**Estimate M_e when *K*=1**		**Estimate M_e when *k*=3**	
		$\bar{e}_s^{(1)}$	$\bar{e}_s^{(2)}$	$(M_e)_3$	**Confidence Interval 99%**	$(M_e)_3$	**Confidence Interval 99%**
1	(V_1,V_2,V_3,V_4)	15.75	7.00	6.760	(3.740, 9.780)	6.760	(3.740, 9.780)
2	(V_1,V_2,V_4,V_5)	15.25	7.25	7.000	(3.980, 10.020)	7.000	(3.980, 10.020)
3	(V_2,V_3,V_4,V_5)	15.00	7.75	7.483	(4.463, 10.503)	7.483	(4.463, 10.503)
4	(V_2,V_3,V_5,V_6)	14.00	7.75	7.481	(4.461, 10.501)	7.481	(4.461, 10.501)
5	(V_3,V_4,V_5,V_7)	15.75	8.50	8.208	(5.188, 11.228)	8.208	(5.188, 11.228)
6	(V_4,V_5,V_6,V_8)	15.50	7.75	7.484	(4.464, 10.504)	7.484	(4.464, 10.504)
7	(V_5,V_6,V_7,V_9)	15.50	7.75	7.484	(4.464, 10.504)	7.484	(4.464, 10.504)
8	(V_6,V_7,V_8,V_9)	16.25	7.00	6.760	(3.740, 9.781)	6.760	(3.740, 9.781)

Table 12 : Sample Based Estimates of Mean Edge in G_2 of Fig. 2 (Related to Table 4)

Note that the computation of 99% confidence interval is according to formula $\left[M_e - 3\sqrt{V(M_e)},\ M_e + 3\sqrt{V(M_e)}\right]$ for values of k= 1 and 3.

6. Discussion

Table 4 presents eight samples, each of four vertices along with the description of edge lengths. Table 5 and 6 are FNEM used to generate Table 4 using node sampling procedure of section 1.1. The Table 7 contains the population-wise details of vertices, linked edges, lengths and average length of each vertex. Table 8 presents the computation of population parameters. The class of estimators M_e contains an unbiased estimator at the choice $k = k_1$ = 2.001256 and there exist only one such real root to satisfy the cubic equation (3.7) [see Table 9]. Minimum mean squared error is found when $k = k_1^{'}$ = 2.026643, satisfying the cubic equation (4.2) with the existence of only one real root [see Table 10]. It is observed that at the optimum value $k^{'}$ = 2.026643,the bias is very small which turns out to explore an almost unbiased minimum variance estimator in the class for estimating mean edge length [see Table 11]. Sample based estimate M_e, for k=1 & k=3, is computed over eight random samples, drawn from population and shown in Table 12. This reveals that the estimate of true length $\bar{e}^{(2)}$ lies in the 99% confidence intervals. There is high chance to get a best estimate of mean edge length of the planer graph population, because the unbiased estimator (k_1= 2.001256) and optimal estimator ($k_1^{'}$=2.026643) both, in the class, are obtainable in the range $1 \le k \le 3$ which generates a sub-class of efficient estimators in the proposed M_e. The sample based estimates are very close to the true estimate $\bar{e}^{(2)}$, when k=1 and 3,as shown in Table 12.

7. Conclusions

Sampling methodology under a graphical population is taken into account and a new sampling technique "Node Sampling Procedure" is designed. A class of estimation strategies is proposed which is found affective to estimate the average length of an edge of planer-graph population .Optimum estimator in the class is obtained and its properties are shown. There are atmost three possible values for which the unbiased estimator could be obtained in the class and one of them is shown. Moreover, the class may have atmost three optimum estimators also, the best would be that having the least bias. One such estimator is obtained on considered data. Node Sampling Procedure facilitates to estimate the mean edge length of planer graph population. The sample based estimates are found closed to the true values. Within range $1 \le k \le 3$, almost unbiased minimum variance estimators are available in the proposed class. Most of the sample estimates depict the true length $\bar{e}^{(2)}$ within the 99% confidence interval specially when k=1 and k=3 are used.

References

1. Aleksandrov, L., Djidjev, H., Guo, H.and Maheshwari, A. (2007). Regular Papers: Partitioning planar graphs with costs and weights, Journal of Experimental Algorithmic (JEA), 11, p. 102-113.
2. Cochran, W.G. (2005). Sampling Techniques, John Wiley and Sons, Fifth Edition, New York.

3. Deo, N. (2001). Graph theory with application to Engineering and Computer Science, Prentice-Hall, Eastern Economy Edition, New Delhi, India.
4. Frederickson, G. N. (1988). Planar graph decomposition and all pairs shortest paths, Journal of the ACM (JACM), p. 162-204.
5. Gazit, H.and Reif, J. (1990). A randomized parallel algorithm for planar graphs isomorphism, Proceedings of the second annual ACM symposium on Parallel algorithms and architectures, SPAA 90, p. 210-219.
6. Grigni, M., Koutsoupias E. and Papadimitriou C. (1995). An approximation scheme for planar graph TSP, 36th Annual Symposium on Foundations of Computer Science (FOCS 95), p. 640-649.
7. Osthus, D., Promel, H. J. and Taraz, A. (2003). On random planar graphs, the number of planar graphs and their triangulations, Journal of Combinatorial Theory (B), 88, p. 119-134.
8. Parson, T. D. (1971). On Planner Graphs, Am. Math. Monthly, 78, No.2, p.176-178.
9. Shih, W. K., Wu S. and Kuo Y.S. (1990). Unifying maximum cut and minimum cut of a planar graph, IEEE Transactions on Computers, p. 694-697.
10. Shukla, D. Singh, V. K. and Singh, G. N. (1991). On the use of transformation in Factor -Type estimator, Metron, 49, 1-4, p. 349-361.
11. Shukla, D.(2002). F-T Estimator under two-phase sampling, Metron, 49, 97-106.
12. Singh, D. and Choudhary, F. S. (1986). Theory and Analysis of Sample Survey Design, Wiley Eastern Limited, New Delhi.
13. Singh, V. K. and Shukla, D. (1987). One parameter family of Factor-Type ratio estimator, Metron, 45, 1-2, p. 273-283.
14. Singh, V. K. and Shukla, D. (1993). An efficient one parameter family of Factor-Type estimators in sample Surveys, Metron, 51, 1-2, p. 139-159.
15. Sukhatme, P.V., Sukhatme, B.V., Sukhatme, S. and Ashok, C. (1984). Sampling Theory of Surveys with Applications, Iowa State University Press, I.S.A.S. Publication, New Delhi.