Small-world in Author Community of Biotechnology: A Study

Tejashwini B.
Chief Librarian
T John College, Gottigere
Bengaluru -560 083
e-mail: teja.buchi@gmail.com

Mallinath Kumbar
Professor and Chairman
Department of Studies in Library and Information Science
University of Mysore, Manasagangothri
Mysuru-570 006
e-mail: mallinathk@yahoo.com

Abstract- *The present research emphasizes the small world phenomenon in author community of Biotechnology. The author-community data for the study has been extracted from Web of Knowledge for tenure of 2003-2012 i.e. ten years. The data was then analyzed using scientometirc and social network analysis criteria's. The results show 94.52% of articles were written by Collaborative authors and 5.47% articles were written by solo authors. Puhler, A stands first in the list of prolific authors of Biotechnology with highest number of articles 62(0.34%) and SNA results also shows that Puhler is highly collaborative. The Characteristic path length of Biotechnology author community is 2.49 which supports small world phenomenon.*

Keywords: Scientometrics, Social network analysis, Small-world phenomenon, Author Community, Biotechnology.

1.Introduction

"It's a Small World" as the phrase says the World or earth is smaller in many terms i.e. Physical distance has been over ruled by Information telecommunications in the Digital era. The Scientist in earlier centuries had huge issue of communication of information with one another, but today scientists are able to communicate information or query with one another with the help of ICT. The scientists working in different fields, subject, and country can easily collaborate for betterment of research and results. Biotechnology is one of the leading areas of research in science, where technology is applied to modify, generate and alter genetic information of species for its betterment. In Biotechnology research community there are scientists from different areas from technology to physical and natural science.

The scholarly communication between scientists and researchers are witnessed through collaborative research article in journals, patents, conference proceedings, Short communications, etc. The study focuses on ties or link between researchers & Scientists through Social network analysis of author collaboration. The researchers/ scientists are said collaborative if they have co-authored a research article. The collaboration ties reveal the strength and weakness of collaboration, which is analysed through degree of collaboration, betweens, closeness and eigenvector.

2. Small-world phenomenon

Stanley Milgram, a social psychologist conducted an experiment called "The small World problem" during 1967 in Harvard University. The experiment was developed out of desire to learn more about the probability that two randomly selected people would know each other (Travers and Milgram, 1967). The experiment was done by randomly selecting people of USA to deliver information pocket or letter to the target person, though 232 of 296 letters never reached the target and 64 of the letters eventually reached the target contact. The average path length fell around five and half or six. Hence the researchers concluded that people in USA are separated by about six people on average. Although Milgram himself never used the phrase "Six degree of separation" these findings are likely to have contributed to its widespread acceptance.(Barabasi, 2003)

Small –world phenomenon is phenomenon which explains that any two individuals in the network are likely to be connected through a short sequence of intermediate acquaintances similarly small world network satisfies two properties according to Watts & Strogatz i.e.

- Small average shortest path (Global)
- High clustering co-efficient (Local)

The small world phenomenon is tested for present study of author community of biotechnology, by calculating characteristic path length of vertices of socio-graph of collaborative author network (Watts& Strogatz, 1998).

3. Objectives of the study

The present study was carried out in view of following objectives:

- To study the multi-authorship trend in Biotechnology.
- To study degree of collaboration in Biotechnology author community
- To study prolific authors in Biotechnology community.
- To study the Social network analysis of author collaboration.
- To study the small world phenomenon in Biotechnology author community.

4. Methodology

The data for the present study was extracted from web of science from 2003-2012 as time constraint, Ten journals pertaining to Biotechnology has been selected which are indexed in Web of Knowledge, which resulted in 18623 full text articles. The data from journals was analyzed in accordance with degree of collaboration, metrics of SNA i.e. degree of closeness, betweeness, eigenvector has been applied to understand the patterns and characteristic of Biotechnologist community for a detailed analysis. The authors have used UCINET commercial software and NetDraw for visualization of network. The results from author collaboration network were further analyzed to understand the small-world phenomenon in Biotechnology author community.

Table -1
List of Journals

Sl. no	Name of Journal	Name of Publisher	Place of publication	Impact factor	Frequency
1.	Food Biotechnology	Taylor & Francis	Philadelphia	0.52	4
2.	Animal Biotechnology	Taylor & Francis	Philadelphia	0.92	4
3.	Food Technology & Biotechnology	University of Zagreb	Croatia	1.19	4
4.	Bioscience, Biotechnology and Biochemistry	Japan Society for Biosciences	Japan	1.27	12
5.	Biotechnology and Bio-process Engineering	The Korean Society for Biotechnology and Bioengineering	Korea	1.27	6
6.	Journal of Industrial Microbiology & Technology	Society of Industrial Microbiology	Fairfax, VA	1.8	12
7.	Journal of Chemical Technology & Biotechnology	Periodicals Service	USA	2.16	12
8.	Journal of Biotechnology	Science Direct	USA	3.04	18
9.	Current Opinion in Biotechnology	Science Direct	Philadelphia USA	7.71	6
10.	Biotechnology Advances	Science Direct	USA	9.64	6

(Source: Web of science, accessed on 12/12/2014)

5. Data analysis and interpretation

The data collected from web of science was analysed using MS-excel, UCINET and NetDraw and presented in the form of tables.

5.1Degree of collaboration

Data was analysed to identify the degree of collaboration various methods have been proposed to calculate the degree of research collaboration. Here, in this study the formula proposed by Subramanyam (1983) has been used.

The degree of collaboration $c = nm/n_m + n_s$
Where, C = Degree of collaboration in a discipline
Nm = number of multi-authored papers in the discipline
Ns = number of single-authored papers in the discipline
Using this formula the degree of collaboration for biotechnology literature, this identifies the closeness of authors in collaboration

Table-2
Degree of collaboration in Biotechnology literature

Sl. No	Years	Single Authorship	Multi Authorship	Total	Degree of collaboration
1	2003	82 (8.04%)	1143 (6.49%)	1225 (6.57%)	0.93
2	2004	64 (6.28%)	1148 (6.52%)	1212 (6.50%)	0.94
3	2005	155 (15.21%)	1737 (9.86%)	1892 (10.15%)	0.91
4	2006	98 (9.61%)	1481 (8.41%)	1579 (8.47%)	0.93
5	2007	71 (6.96%)	1901 (10.79%)	1972 (10.58%)	0.96
6	2008	61 (5.98%)	1662 (9.44%)	1723 (9.25%)	0.96
7	2009	71 (6.96%)	1830 (10.39%)	1901 (10.20%)	0.96
8	2010	186 (18.25%)	2878 (16.34%)	3064 (16.45%)	0.93
9	2011	132 (12.95%)	1986 (11.28%)	2118 (11.37%)	0.93
10	2012	99 (9.71%)	1838 (10.44%)	1937 (10.40%)	0.94
Total		1019	17604	18623	0.94
Percent age		5.47	94.52	100	

Table-2 depicts degree of collaboration among authors by using Subramanyam's formula. The table shows the single authorship versus multiple authorship articles during 2003-2012; in 2003 single authored articles are 82 versus multi-authored articles 1143 of total output 1225 with degree of collaboration 0.93, followed by single authored articles 64 versus multi-authored articles 1148 total output of 1212 articles with degree of collaboration rate 0.94 in 2004 and 155 single authored articles versus 1737 multi-authored articles of 1892 total output with degree of collaboration rate 0.91 in 2005. The number of single authored articles was 98 versus multi-authored articles 1481 of 1579 total output with degree of collaboration rate 0.93 in 2006, followed by single authored 71 articles versus multi-authored 1901 articles of 1972 total output, along with degree of collaboration rate 0.96 in 2007 and in 2008 single authored articles was 61 versus multi-authored 1662 articles out of 1723 total output with degree of collaboration rate 0.96. Out of 1901 total articles 71 was single authored versus 1830 multi-authored articles with degree of collaboration rate 0.96 in 2009, followed by 186 single authored articles versus multi-authored 2878 articles of 3064 total output with degree of collaboration rate 0.93 in 2010 and in 2011 the total output articles 2118 with single authored 132 articles versus 1986 multi-authored articles with degree of collaboration rate 0.93, in 2012 the total output articles 1937 with single authored 99 articles while multi-authored 1838 articles along with degree of collaboration rate 0.94. The total single authored articles during 2003-2012 was 1019 i.e. 5.47 % while multi-authored articles 17604 with 94.52% of total output of 18623 with degree of collaboration 0.94.

5.2 Prolific authors

Authors Data was analyzed to identify the prolific authors of Biotechnology literature in 2003-2012 years. The top most authors has published highest number articles during 2003-2012, the table is generated with descending order of number of article published by each authors.

Table-3
Prolific authors of Biotechnology literature

Sl. No	Authors	No. Of Articles	% for 18623	Cumulative Percentage	Sl. No	Authors	No. of Articles	% For 18623	Cumulative Percentage
1	Puhler A	62	0.34	0.34	21	Matsushita K	32	0.17	4.54
2	Takahashi K	51	0.28	0.62	22	Kim SW	32	0.17	4.71
3	Kimura T	50	0.27	0.89	23	Lee JH	30	0.16	4.87
4	Nakamura Y	44	0.24	1.13	24	Park JK	30	0.16	5.03
5	Abe K	42	0.23	1.36	25	Nielsen J	30	0.16	5.19
6	Kimura M	41	0.22	1.58	26	Levin Re	28	0.15	5.34
7	Mizuno T	40	0.22	1.8	27	Pandey A	28	0.15	5.49
8	Watanabe T	40	0.22	2.02	28	Goesmann A	28	0.15	5.64
9	Yamauchi S	40	0.22	2.24	29	Tauch A	26	0.14	5.78
10	Yoshida T	40	0.22	2.46	30	Neubauer P	25	0.13	5.91
11	Kimura Y	39	0.21	2.67	31	Kim JH	24	0.13	6.04
12	Suzuki Y	38	0.2	2.87	32	Kim SK	23	0.12	6.16
13	Yamada K	38	0.2	3.07	33	Kalinowski J	23	0.12	6.28
14	Sato M	36	0.19	3.26	34	Gasser RB	21	0.11	6.39
15	Watanabe K	36	0.19	3.45	35	Fava F	21	0.11	6.5
16	Suzuki T	35	0.19	3.64	36	Olsson L	21	0.11	6.61
17	Yamamoto K	35	0.19	3.83	37	Soccol CR	20	0.11	6.72
18	Murata Y	34	0.18	4.01	38	Kim SW	20	0.11	6.83
19	Watanabe H	34	0.18	4.19	39	Mackova M	20	0.11	6.94
20	Yoshida M	33	0.18	4.37	40	Seo JH	20	0.11	7.05

Table-3 Shows the prolific authors of Biotechnology literature during 2003-2012, top 40 authors out of 18623 publications has been listed here with highest number of articles published by Puhler A with 62 (0.34%), followed by Takahashi K with 51(0.28%) articles, in third position is Kimura T with 50 (0.27%) articles, at four is Nakamura Y with 44 (0.24%) articles, followed Abe K with 42 (0.23%) articles at fifth position. Kimura M stands sixth with 41 (0.22%) articles, followed by Mizuno T with 40 (0.22%) articles at seventh position, Watanabe T 40 (0.22%) articles, Yamauchi S with 40 (0.22%) articles and Yoshida T with 40 (0.22%) articles, followed by Kimura Y with 39 (0.21%) articles at eight position. Suzuki Y 38 (0.2%) articles are at ninth position along with Yamada K 38 (0.2%) and at tenth position is Sato M with 36 (0.19 %) articles along with Watanabe K 36 (0.19 %) articles and at forty is Seo J H with 20 (0.11%) articles.

5.3 Social Network Analysis of Author Collaboration

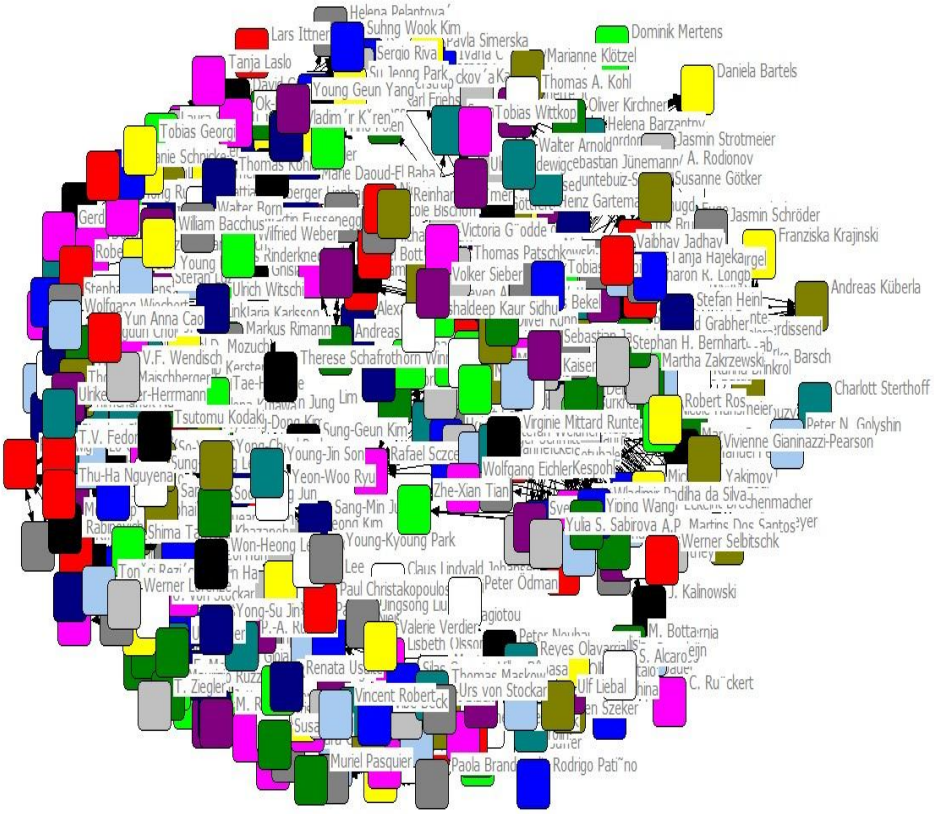

Figure -1 Author Collaboration of Biotechnology Literature

Figure-1 shows the sociograph of Author collaboration of Biotechnology literature during 2003-2012, the sociograph shows the co-authorship network and ties of actors i.e. Authors through research collaboration in Biotechnology literature. The author's data from 18632 articles has been filtered by investigator by selecting authors of biotechnology, who have co-authored at least 2 research articles together. Based on this we obtained 936 authors from ten journals who have collaborataed during 2003-2012.

The socio-graph of author collaboration is undirected, small world phenomenon is analysed using characteristic path length between the nodes or authors. The graph is connected if there is a path of finite length between any two nodes. The Maximum distance i.e. path length between any two connected vertices of a graph based on finite diameter or path length, based on this the network is said to be connected or not connected.

5.3.1Degree Centrality

Degree centrality is defined as the ratio of the number of neighbors of a vertex with the total number of neighbors possible. Mathematically,

$$\text{Degree Centrality} = \frac{k}{N - 1}$$

Where k is the degree of the vertex, and N is the total number of nodes in the network. The variance of the distribution of degree centrality in a network gives us the centralization of the network.

5.3.2 Betweenness Centrality

The degree of a node is not the only measure of the importance of a node in the network, and this centrality measure addresses this fact. This concept was introduced by Linton Freeman. In his conception, vertices that have a high probability of occurring on a randomly chosen shortest path between two no desire said to have high betweenness centrality.
Formally, centrality of a vertex v is defined as the summation of the geo-disc path between any two nodes s and t via v, expressed as a fraction of the total number of geodesic paths between s and t.

Mathematically, $g(v) = \sum_{s \neq v \neq t} \frac{\sigma st(v)}{\sigma st}$

5.3.3 Closeness Centrality

Closeness centrality is used to find central vertices. It gives higher values to more central vertices. Closeness centrality of a node x, is denoted by Cres(x) and is calculated as follows
Cres(x) = $\sum_{y \in U, y \neq x}^{N-1} d(x,y)$
Where d(x, y) is the geodesic distance between node x and node y. U is the set of all nodes and N is the number of nodes in the network. The closeness value is therefore the inverse of the average distance between x and other nodes (d).

5.3.4 Eigenvector Centrality

The value ƛ is an eigenvalue of matrix A if there exists a non-zero vector x, such that Ax =ƛx. Vector x is an eigenvector of matrix A The largest eigenvalue is called the principal eigenvalue The corresponding eigenvector is the principal eigenvector Corresponds to the direction of maxi mum change Eigenvector centrality is another measure of influence of a node in a network. It assigns relative scores to all nodes in the network based on the concept that connections to high-scoring nodes contribute more to the score of the node in question than equal connections to low-scoring nodes.

Table-4
Centrality Measures Author collaboration in Biotechnology Literature

Author Rank	Author Name	Degree	Closeness	Betweennes s	Eigenvecto r
1.	Alfred Puhler	28.17	0.44	6.56	-46.11
2.	Takahashi K	15.70	0.44	2.93	-29.52
3.	Kimura T	14.31	0.44	0.73	-32.98
4.	Gasser R B	12.93	0.44	1.30	-24.11
5.	Mantzavinos D	10.39	0.44	0.63	-22.52
6.	Chen J	10.16	0.44	0.36	-23.36
7.	Bennett GN	8.08	0.44	3.98	-4.61
8.	Baltz RH	8.08	0.44	0.32	-18.37
9.	Soccol CR	7.62	0.24	0.42	0
10.	Shetty K	6.92	0.44	0.12	-19.34
11.	Pandey A	6.69	0.24	0.33	0
12.	Levin RE	6.46	0.44	0.13	-17.84

13.	Dundar M	6.46	0.44	0.51	-14.99
14.	Ghasemi Y	6.46	0.44	0.11	-17.20
15.	Kim SW	6.00	0.44	0.06	-18.66
16.	Lee JH	5.77	0.44	0.86	-13.08
17.	Li N	5.77	0.24	0.21	0
18.	Reed KM	5.31	0.44	0.17	-14.61
19.	Chisti Y	5.08	0.44	0.03	-13.52

Table-4 depicts centrality measures of Author Collaboration of Biotechnology literature of ten Journals together for 2003-2012. The Social network analysis reveals that Alfred Puhler stands first in author collaboration with 28.17 degree of centrality, followed by 0.44 closeness centrality along with 6.56 betweeness centrality and eigenvector of-46.11.Takahashi K stands second on the ranking in collaboration with degree of centrality 15.70, followed by 0.44 closeness centrality along with betweeness centrality 2.93and eigenvector -29.52 of on the list at twentieth position is Chisti Y with 5.08 degree of centrality, followed by 0.44 closeness centrality along with 0.03betweeness centrality and eigenvector -13.52.The core members of author collaboration network of Biotechnology community are Puhler A, Takahashi K, Kimura T and Gasser R B.

Table-5
Descriptive Statistics of Centrality Measures Author collaboration of Biotechnology literature

	Degree	**Closeness**	**Betweenness**	**Eigenvector**
Mean	2.05	0.33	0.07	-3.43
Std Dev	2.49	0.09	0.58	5.85
Sum	890.06	145.64	32.96	-1489.83
Variance	6.24	0.01	0.34	34.29
Eigenvector Normalized	67.33	7.29	12.34	141.42
Minimum	0.00	0.23	-0.00	-49.81
Maximum	28.17	0.44	8.89	0.00
Number of Observations	936.00	936.00	936.00	936.00

Table-5 shows descriptive statistics of centrality measures of Author collaboration of Biotechnology literature with mean value of degree of centrality is 2.05, followed by closeness centrality mean value of 0.33, betweens centrality mean value 0.77 and -3.43 mean value of eigenvector. The standard deviation values of four centrality measures are calculated i.e. 2.49 degree of centrality standard deviation value, followed by 0.09 Std Dev closeness centrality, betweenness centrality Std Dev value 0.58 and eigenvector Std dev 5.85. The sum value of degree centrality is 890.06, followed by closeness centrality sum value 145.64, betweenness centrality sum value is 32.96 and eigenvector sum value is -1489.83. The variance degree centrality is 6.24, followed by closeness centrality variance 0.01, betweens centrality variance 0.34 and eigenvector variance value 34.29. The minimum degree centrality value is 0.00, followed by closeness centrality minimum value 0.23, betweenness centrality minimum value -0.00 and eigenvector minimum value -49.81. The Number of observations i.e. 936 prolific authors has been observed in Social network analysis of Author collaboration network.

5.3.5 Characteristic Path Length of a Network

The characteristic path length (L) of a network is the shortest path length between two nodes averaged over all pairs of nodes and is given by

$$L = \frac{\in i \in jLij}{N(N-1)}$$

Where Li, j is the shortest path length between i^{th} node and j^{th} node. Higher characteristic path length implies network is almost in liner chain and lower characteristic path length shows the network is in compact form. A path in a graph is a single vertex or an ordered list of distinct vertices. Similarly characteristic path length of a network L of a network is shortest path length between nodded averaged over all pairs of nodes.

The data from table-4 and Table-5 has been applied to calculate characteristic path length of author network of biotechnology community. The characteristic path length is 2.49 of author network; this means that the Biotechnology co-authorship network is small world, the Biotechnology co-authorship network has short characteristic path length, indicating that the authors from different groups are well connected.

6. Findings

Based on the analysis following findings have been drawn.

- 94.52% of articles are published by multi-authorship and 5.47% articles are published by solo authors
- Highest number of articles are published in 2010 with 3064 (16.45%) and lowest number of articles are published in 2004 with 1212 (6.5%)
- Puhler, A stands first in publishing highest number of articles with 62(0.34%)
- 936 authors have co-authored minimum 2 research articles in biotechnology journals during 2003-2012.
- Author data was analysed for centrality measures, Puhler, A stands first among all authors in Degree of centrality with 28.17 and 6.56 betweeness centrality.
- The characteristic path length of socio –graph of author data is 2.56 which indicated that authors of biotechnology are well connected and small world phenomenon of Stanley milligram is true to this data.

7. Conclusion

The Biotechnology author community is very collaborative through the number of articles by Multi-authors i.e. 94.52%, social network analysis of author network gives a socio-graph which is densely covered with nodal actors i.e. authors. The small-world phenomenon of Milgram is also applicable to author network of Biotechnology community. The socio-graph and descriptive statistics reveal that characteristic path length between authors in author network is 2.49 which proves that authors in biotechnology are well connected with each other and social network analysis & proves that small world phenomenon is applicable to Authors of biotechnology community.

References

1. Barabasi, Albert-Laszio (2003) “Linked: How everything is connected to everything else and What it means for business, science and Everyday life, New York: Plume.

2. Barnes, J.A. (1983). "Graph theory in network analysis" *Social Networks,* 5: 235-244.
3. Borgatti, Steven, Martin Everett, and Linton Freeman. (1992). UCINET IV Version 1.0 User's Guide. Columbia, SC: Analytic Technologies.
4. D'Andrea, Alessia et al. (2009). "An Overview of Methods for Virtual Social Network Analysis". In Abraham, Ajith et al. *Computational Social Network Analysis: Trends, Tools and Research Advances*. Springer. ISBN 978-1-84882-228-3.
5. Marsden, Peter and Nan Li (1982). Social structure and network analysis. Beverly Hills: Sage.
6. Milgram, Stanely (1967) "The small-world problem" Psychology Today. Ziff-Davis Publishing Company.
7. Otte, Evelien; Rousseau, Ronald (2002). "Social network analysis: a powerful strategy, also for the information sciences". *Journal of Information Science* **28**: 441–453. doi:10.1177/016555150202800601. *Retrieved 2015-03-23*.
8. Scott, John. (1991) Social Network Analysis: A Handbook. Newbury Park, CA: Sage Publications.
9. Travers, J and Milgram, S. (1969). N Experimental study of the small-world problem. *Sociometry*, 32(4): 425-443.
10. Watts, D.J. and Strogatz, S.H. (1998). Collective dynamics of small –world networks. *Nature,* 393(6684): 440-442.