Sub-field Visualization: A Multidimensional Analysis of Web 2.0 Authors

Edward Benoit, III
School of Information Studies
University of Wisconsin-Milwaukee
eabenoit@uwm.edu

Scholars often consider Library and Information Science as an interdisciplinary field. While most in the field focus on Information Science specifically, many enter the field from outside disciplines including the sciences, the social sciences, and the humanities. Similarly, research within Library and Information Science has vast implications throughout academic disciplines. Although this has always been the case, recent emerging trends within technological developments increased the interdisciplinary tendencies. Research and use of new Internet based applications comes from many areas, and its implications reach to equal varieties.

Internet technologies continuously grow at a rapid rate. The past decade saw the emergence of Web 2.0, a change in the ways the internet applied to its users. Applications such as Facebook, Flickr, Wikipedia, and blogging software illustrate just some of the uses of Web 2.0 technologies. Since Web 2.0 applications range across disciplines, could their development and research be interdisciplinary? Furthermore, what relationships connect researchers from this field?

Similarly, multidimensional scaling (MDS) as an analytical tool began increasing in popularity over the past decades. Used throughout the sciences and social sciences, MDS allows researchers a unique visualization technique for determining relationships. Although widespread in use, additional MDS applications exist. Multidimensional scaling could explore the relationships between authors within a given field. This case study tests such an application, through an exploration of Web 2.0's most published authors.

Literature Review

The field of psychology developed and initiated the use of multidimensional scaling as a methodology over several decades. One of the earliest works investigated the use of MDS as a comparison technique between two or more psychophysical scales (Young & Householder, 1941). Another study introduced MDS as a solution to the problems of unknown dimensional numbers. As Torgerson states, "In many stimulus domains, however, the dimensions themselves, or even the number of relevant dimensions, are not known. What might appear intuitively to be a single dimension may in fact be a complex of several" (Torgerson, 1952, p. 401). Further development of MDS culminated in Guttman's application of the method to represent similarities within a coordinated space (1968).

Although originated in psychology, MDS expanded through the sciences and social sciences as a method for the visualization of similarities, distances, and relationships. The method translates "proximities" into a coordinated, low-dimensional space, thus allowing user manipulation and analysis. As Zhang summarizes:

> Input data for MDS analysis is usually a measure of proximity (similarity or dissimilarity) of investigated objects in a high dimensional space, while its output is a spatial object configuration in a low dimensional space where users may perceive and analyze the relationships among the displayed objects. It is apparent that in such a MDS display space the more similar two objects, the closer to each other they are, and vice versa (2008, p. 143).

Recent Library and Information Science MDS applications focus on query analysis. These include an analysis of frequently used query terms within a health services setting (Zhang, Wolfram, Wang et al., 2008), and the comparison of sport related queries between term assisted and non-assisted applications (Zhang, Wolfram & Wang, 2009). MDS outside of Library and Information Science vary in both application and field. Pardoe, for example, proposes MDS to assist in grouping college students based on schedule availability (2004). In archaeology, MDS confirms the validity of "late period phases in the Central Mississippi Valley" (Mainfort, 2003, p. 176). The applications continue from social relationships among baboons (Easley, 1990) to tourism research (Fenton, 1988).

Another development within MDS applications is its use as a field exploration tool. Specifically, both existing and emerging academic fields. Biglan compared the similarities between 36 different academic fields as judged by 168 faculty members at the University of Illinois (1973) and 54 faculty members from a small liberal arts college. His analysis found three distinct divisions within academia: the hard science-soft science division; a division based on the application of research; and a division between fields studying animate versus inanimate objects. Another study applied co-citation analysis and MDS

Table 1 Sample Population by Number of Publications (Author [Assigned Number])

5 Publications	**2 Publications**	LANKES, RD (49)
BEER, D (4)	AL-SHAHROUR, F (17)	LARSON, EL (50)
4 Publications	ALLISON, M (18)	LIU, Y (51)
AHARONY, N (1)	ALLOZA, E (19)	LUCKMAN, S (52)
CHEUNG, KH (3)	BAEZA-YATES, R (20)	LUDVIGSSON, J (53)
3 Publications	BECVAR, KM (21)	LUGMAYR, A (54)
BAWDEN, D (2)	BOAST, R (22)	MAJCHRZAK, A (55)
BOULOS, MNK (23)	BUCKLEY, N (24)	MCCLURE, M (56)
COOKE, M (6)	BURNHAM, JF (25)	MEDINA, I (57)
DELLAVALLE, RP (34)	CARBONELL, J (26)	MIKA, P (58)
GUALLAR, J (41)	CHAPMAN, S (27)	MINGUEZ, P (59)
HANBERGER, L (7)	CHAWNER, B (28)	MONTANER, D (60)
HARDEY, M (42)	CHIANG, IP (29)	NGO, CW (61)
HUGHES, B (5)	CHU, HT (30)	SETHI, SK (62)
LI, Q (8)	CHURCHILL, D (31)	SILVERSTEIN, J (63)
NORDFELDT, S (9)	CONESA, A (32)	SO, HJ (64)
PARK, J (10)	DELGADO-LOPEZ-COZAR, E (33)	SRINIVASAN, R (65)
PRECIADO, JC (11)	DOPAZO, J (35)	TORRE, I (66)
SANCHEZ-FIGUEROA, F (12)	EKBERG, J (36)	TOWNSEND, JP (67)
SANDARS, J (13)	ERIKSSON, H (37)	TSAI, CC (68)
SCOTCH, M (14)	FREEMAN, B (38)	WAGNER, C (69)
TIMPKA, T (15)	FURNER, J (39)	WAREHAM, J (70)
TORRES-SALINAS, D (16)	GOETZ, S (40)	WEIKUM, G (71)
	HUANG, YM (43)	WUSTEMAN, J (72)
	JONES, J (44)	XU, C (73)
	JONES, N (45)	YIP, KY (74)
	JOSHI, I (46)	ZHUGE, H (75)
	KIM, S (47)	ZUMER, M (76)
	KIND, T (48)	

for an investigation of the development of management information systems (Culnan, 1986). The analysis found nine groupings and concluded the system development lacked organizational theory. Similar to the current study, exploration of an emerging field, the academic discipline of Urban Studies underwent a MDS analysis in an attempt to define itself (Bowen, Dunn & Kasdan, 2010). The study found an internal, three-dimensional structure in Urban Studies, based on survey data.

Overall, MDS developed over the past century as a visualization and exploratory methodology out of psychological analysis. During the past thirty years, its application spread throughout the social and hard sciences. Despite its widespread nature, MDS receives only limited use as an academic field analytical tool. The few previous studies used either quali-

tative data, such as surveys, or co-citation analysis. Multidimensional scaling's use for discovering the relationships between the most published authors within an emerging field remains an innovative technique, whose results may indicate further applications.

Methodology

The multidimensional scaling analysis requires four sequential stages: sampling, keyword matrix, author similarity matrix, and MDS. Discussed separately below, each of the stages requires the implementation of variable parameters.

Sampling

Identification of the most published authors within the subfield of Web 2.0 used ISI Web of Knowledge's topic search feature. The query "Web 2.0" found 580 publications from 1219 different authors. Internal result analysis identified 76 authors who published more than one article (excluding anonymous authors, conference proceedings/papers, and book reviews). Extracting the authors' names and publishing count created the final sample population. Table 1 lists the authors, publishing count, and assigned number used for tracking authors within the study.

The authors' current academic department or company determined their assigned general research field. Final analysis of the MDS results interpreted clusters based on these research fields. The sample population consisted of the following research fields: Medicine/Health (26.3%), Biomedicine (5.3%), Bioinformatics (13.2%), Library/Information Science (26.3%), Computer Science (19.7%), Education (3.9%), Sociology (1.3%), Communication (2.6%), and Business (1.3%).

Keyword Matrix

The creation of a keyword matrix, representing each authors research profile (not limited to Web 2.0), required the compilation of all published journal articles for each author. Web of Knowledge limited the included articles to those published within ISI indexed journals. An author search within Web of Knowledge produced a comprehensive listing of ISI ranked publications. After the exclusion of conference proceedings/papers and book reviews, an aggregated compilation of each entry's Subject Category and KeyWords Plus (both assigned by ISI) set the research profile for each author. Although some authors provided additional keywords for articles, the study excluded them due to their uncontrolled nature and variability. The creation of a keyword/index term frequency matrix used the aggregated list of terms.

$$\begin{pmatrix} & T_1 & T_2 & \cdots & T_{5411} \\ a_1 & 0 & 0 & ... & 0 \\ a_2 & 0 & 0 & ... & 0 \\ \vdots & \vdots & \vdots & ... & \vdots \\ a_{76} & 0 & 0 & \cdots & 0 \end{pmatrix}$$

Author Similarity Matrix

Constructed using the keyword matrix, the author similarity matrix compares the similarities between authors based on the absolute value of a Pearson's correlation coefficient between authors. Pearson's correlation coefficients apply normalization standards, required for MDS, thus its use rather than non-normalized similarity measurement techniques.

$$\begin{pmatrix} & a_1 & a_2 & \cdots & a_{76} \\ a_1 & 1 & 0.572 & ... & 0.328 \\ a_2 & 0.572 & 1 & ... & 0.585 \\ \vdots & \vdots & \vdots & ... & \vdots \\ a_{76} & 0.328 & 0.585 & \cdots & 1 \end{pmatrix}$$

Multidimensional Scale Analysis

Using the author similarity matrix, a MDS analysis was conducted through the SPSS statistical software. The analysis used the Minkowski interval measure at a power of two, calculated Kruskal stress values, and created a three dimensional model. A hierarchical cluster analysis confirmed the MDS analysis. Additional visualization of the model used the Graphis software platform with the MDS dimensional coordinates. Authors' research field identification (assigned during the sampling stage) added another dimension to the coordinates to explore further relationships between points within clusters.

Results

Initial MDS Model

The three-dimensional model resulted from the MDS analysis, with a stress value of 0.089 and a squared correlation of 0.96. Since the stress value falls under 0.10, the analysis meets goodness of fit measures, indicating the low-dimensional space project faithfully configures to the high-dimensional space. Figure 1 illustrates the MDS findings, and the colorization of points indicates additional research field information. The image demonstrates a strong vertical column of authors from the Medicine/Health, Biomedical, and Bioinformatics. Although this grouping appears obvious, the extension of several of the group's authors into the area populated by Computer Science and Library/Information Science suggests occasional topical overlap.

The initial MDS model (Fig. 1) also illuminates the relationship between computer science and Library and Information Science. While both fields populate the center of the three-dimensional space, a closer examination notes the Library and Information Science authors occur in more condensed pockets.

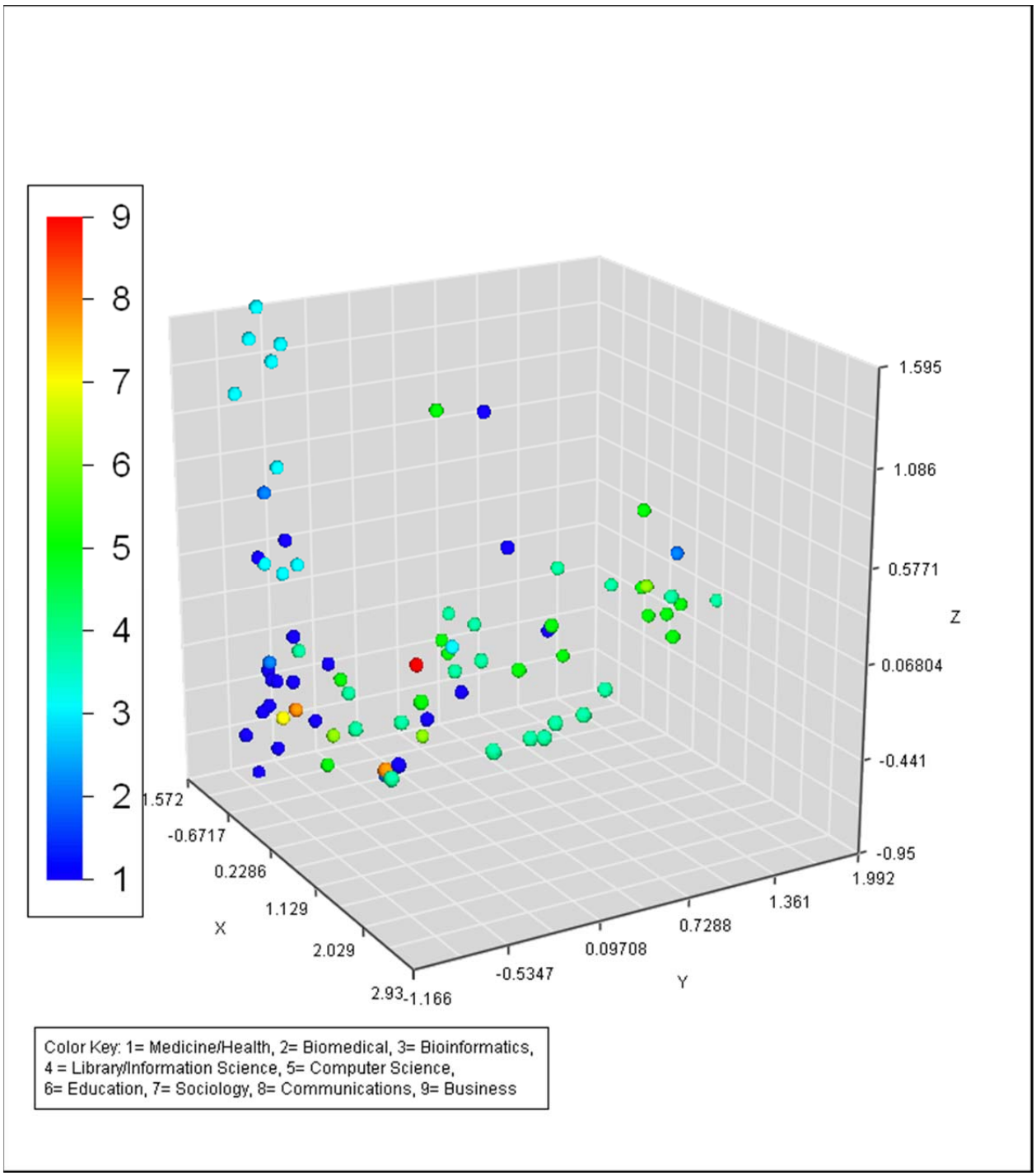

Figure 1 MDS Model with Research Field Colorization

This suggests Computer Science authors are more diverse in their research than Library and Information Science. Figure 2 better highlights these differences.

Finally, the initial model places the remaining fields of Education, Sociology, Communications, and Business within close proximity to the most densly populated region of the scatterplot. These locations still include small differences based on their proximity to the Medical/Biomedical/Bioinformatic column described earlier or the Computer Science/Library and Information Science grouping. Either instance indicates the authors' close relationship

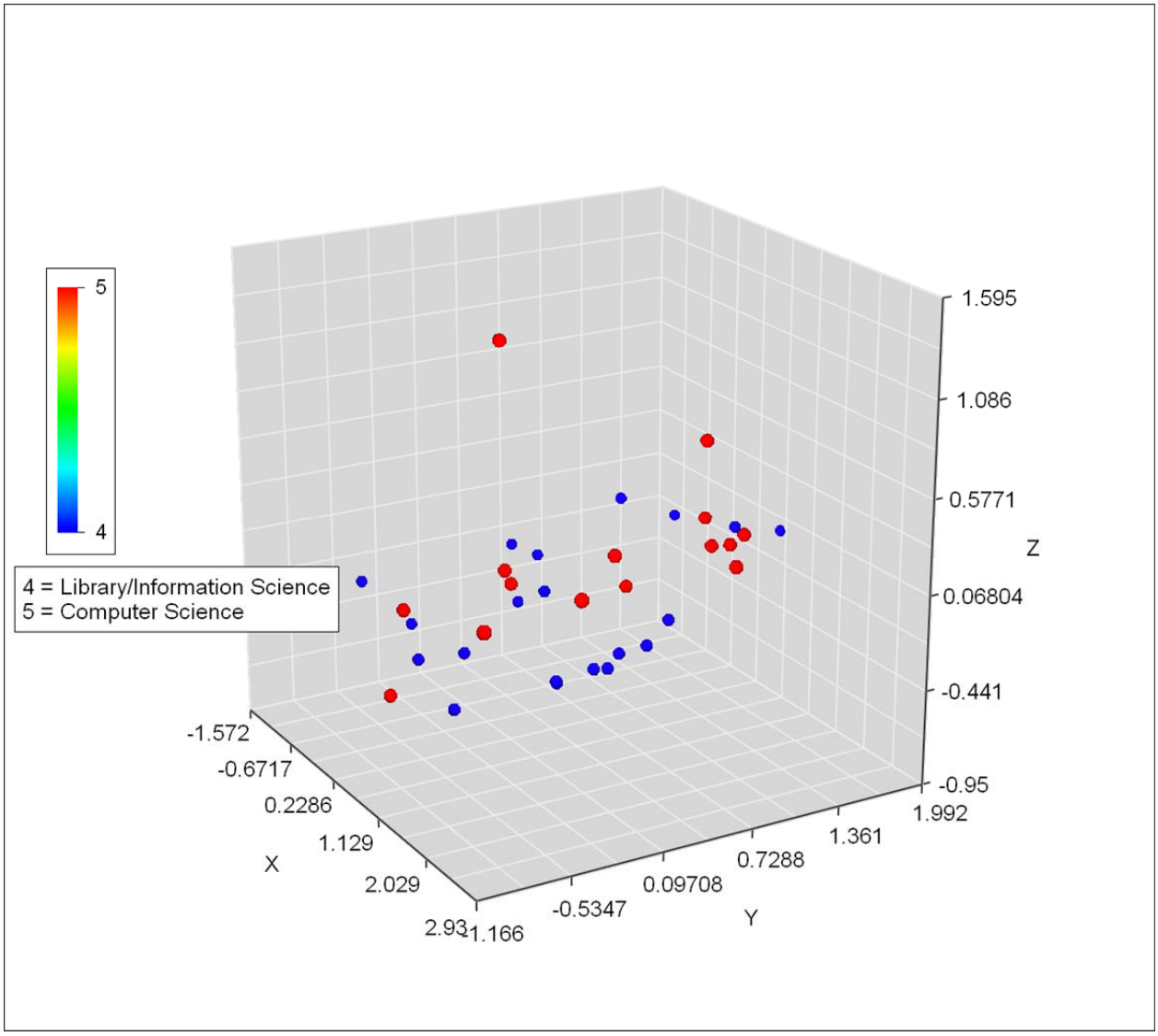

Figure 2 MDS Model with Computer Science and Library/Information Science Colorized

with outside disciplines, also suggesting possible topical overlap between fields.

Hierarchical Cluster Analysis

A second three-dimensional scatter plot using the MDS analysis results and colorized with the hierarchical cluster analysis indicates 18 distinct clusters, ranging between 2 and 13 members. The model displays the hierarchical structure due to the colorization of clusters in numerical order (Fig. 3), whereas clusters similar in color fall close within the hierarchy. Figure 4 outlines and identifies each cluster while Table 2 lists the membership of the groupings.

Cluster Analysis versus Research Field Analysis

A comparison of the cluster analysis model (Fig. 4) and the original MDS model highlighting the authors' research fields (Fig. 1) indicate several interesting anomalies. Although a majority of the clusters fall within one or two closely related fields (such as Biomedical and Medicine/Health), some do not follow this trend. The overlaid model (Fig. 5) shows seven different clusters (C1, C6, C8, C13, C14, C15, and C17) with mixed memberships of unaligned fields. The largest, cluster 1, contains authors from five fields, for example. The existence of clusters with multiple research fields suggests possible t between authors. Additionally, the overlaid model indicates possible subfields within each discipline (e.g. a pediatrics specialty within the Medicine/Health field).

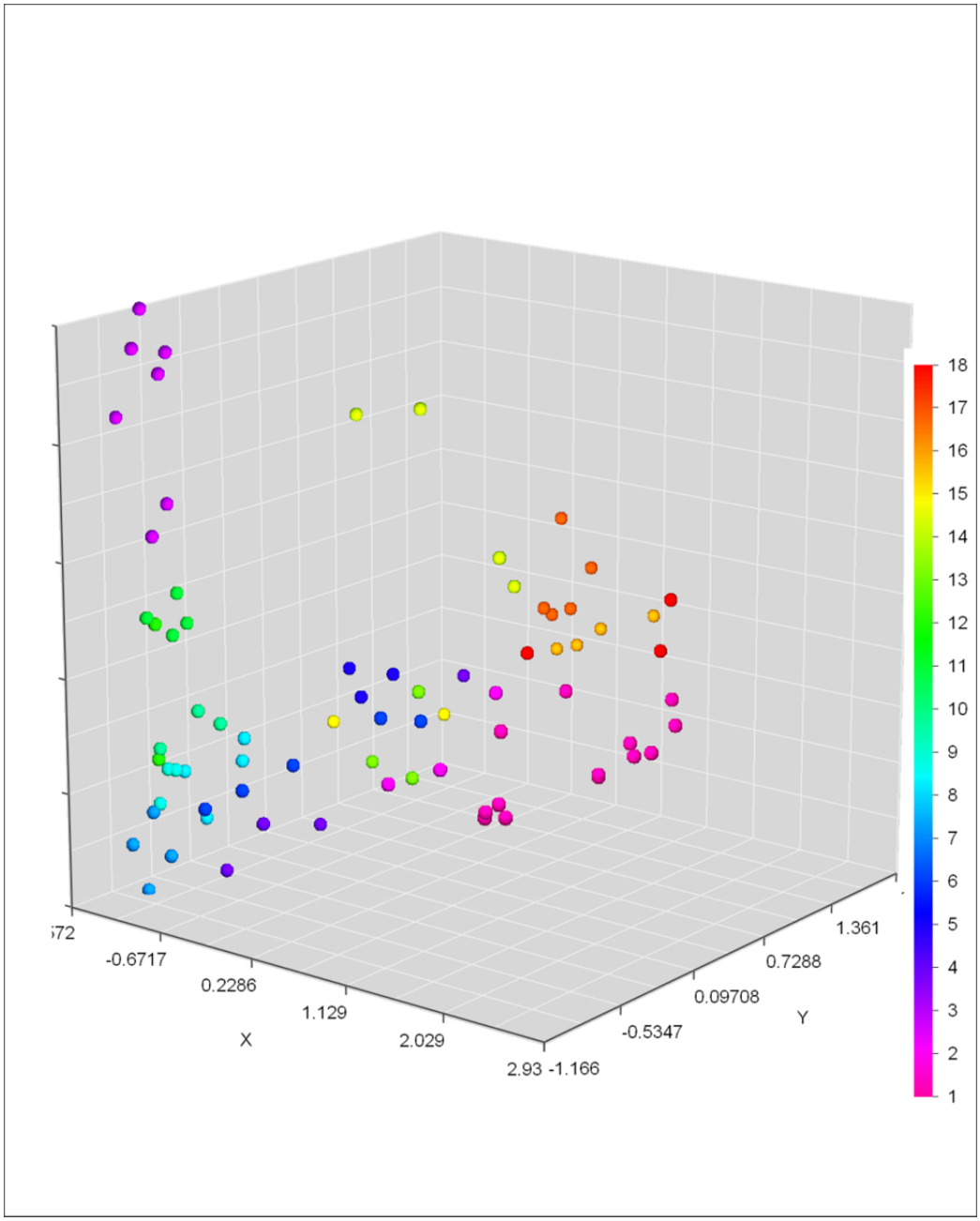

Figure 3 MDS Model with Cluster Colorization

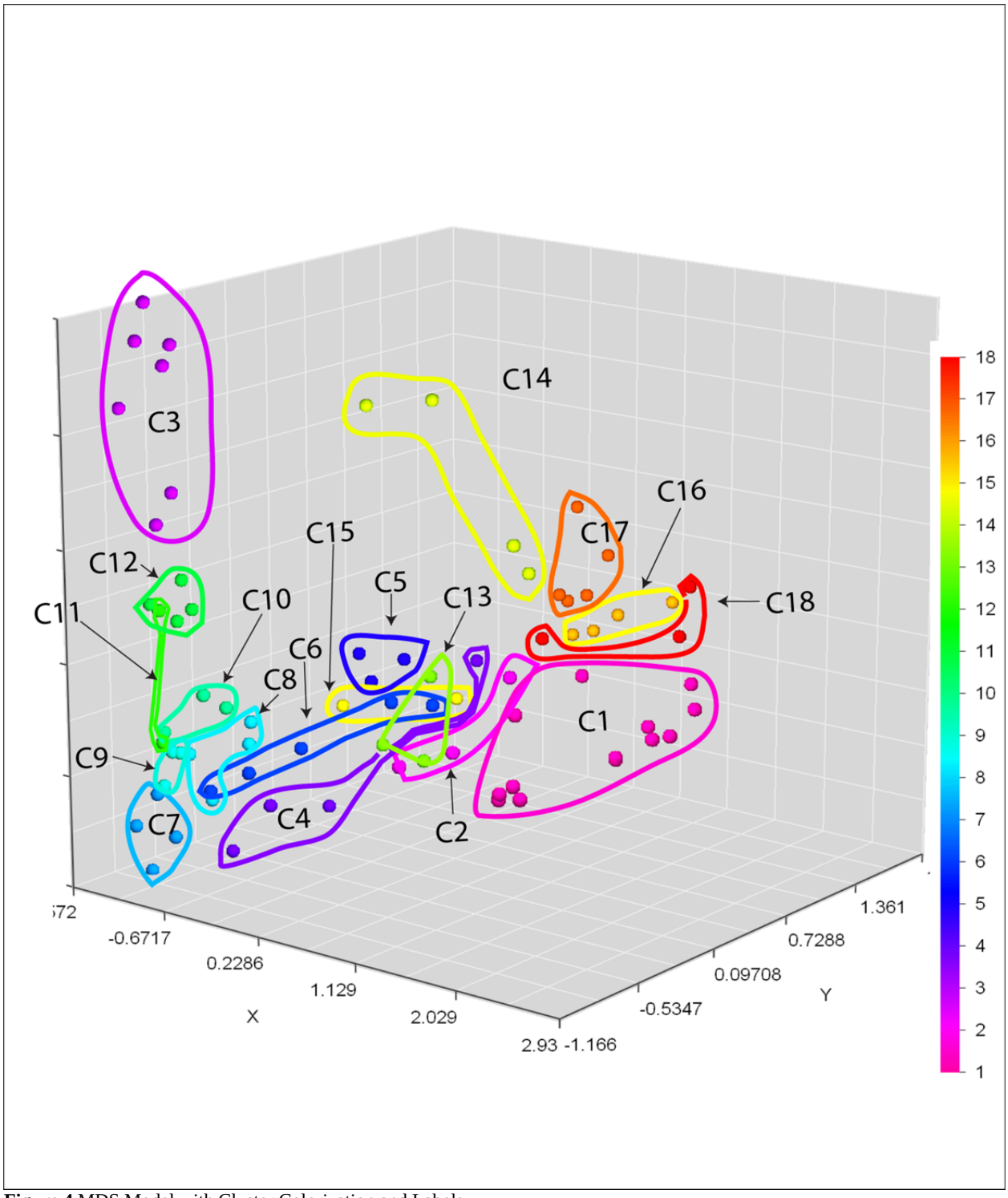

Figure 4 MDS Model with Cluster Colorization and Labels

Table 2 Cluster Identification and Membership by Author

Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5
Bawden, D	Aharony, N	Al-Shahrour, F	Churchill, D	Hughes, B
Torres-Salinas, D	Guallar, J	Alloza, E	Jones, J	Mika, P
Becvar, K	Zumer, M	Conesa, A	So, HJ	Torre, I
Burnham, JF		Dopazo, J	Tsai, CC	
Chawner, B		Minguez, P		
Chu, HT		Montaner, D		
Delgado-Lopez-Cozar, E		Townsend, JP		
Furner, J				
Lankes, RD				
McClure, M				
Silverstein, J				
Wusteman, J				
Xu, C				
Cluster 6	**Cluster 7**	**Cluster 8**	**Cluster 9**	**Cluster 10**
Beer, D	Hanberger, L	Cooke, M	Chapman, S	Buckley, N
Boast, R	Nordfeldt, S	Sandars, J	Hardey, M	Chiang, IP
Luckman, S	Ludvigsson, J	Allison, M	Larson, EL	Joshi, I
Majchrzak, A	Sethi, SK	Freeman, B		
Wareham, J				
Cluster 11	**Cluster 12**	**Cluster 13**	**Cluster 14**	**Cluster 15**
Carbonell, J	Dellavalle, RP	Scotch, M	Cheung, KH	Preciado, JC
Ekberg, J	Medina, I	Timpka, T	Jones, N	Sanchez-Figueroa, F
Goetz, S		Boulos, MNK	Srinivasan, R	
Kind, T			Yip, KY	
Cluster 16	**Cluster 17**	**Cluster 18**		
Park, J	Li, Q	Baeza-Yates, R		
Lugmayr, A	Eriksson, H	Liu, Y		
Weikum, G	Huang, YM	Wagner, C		
Zhuge, H	Kim, S			
	Ngo, CW			

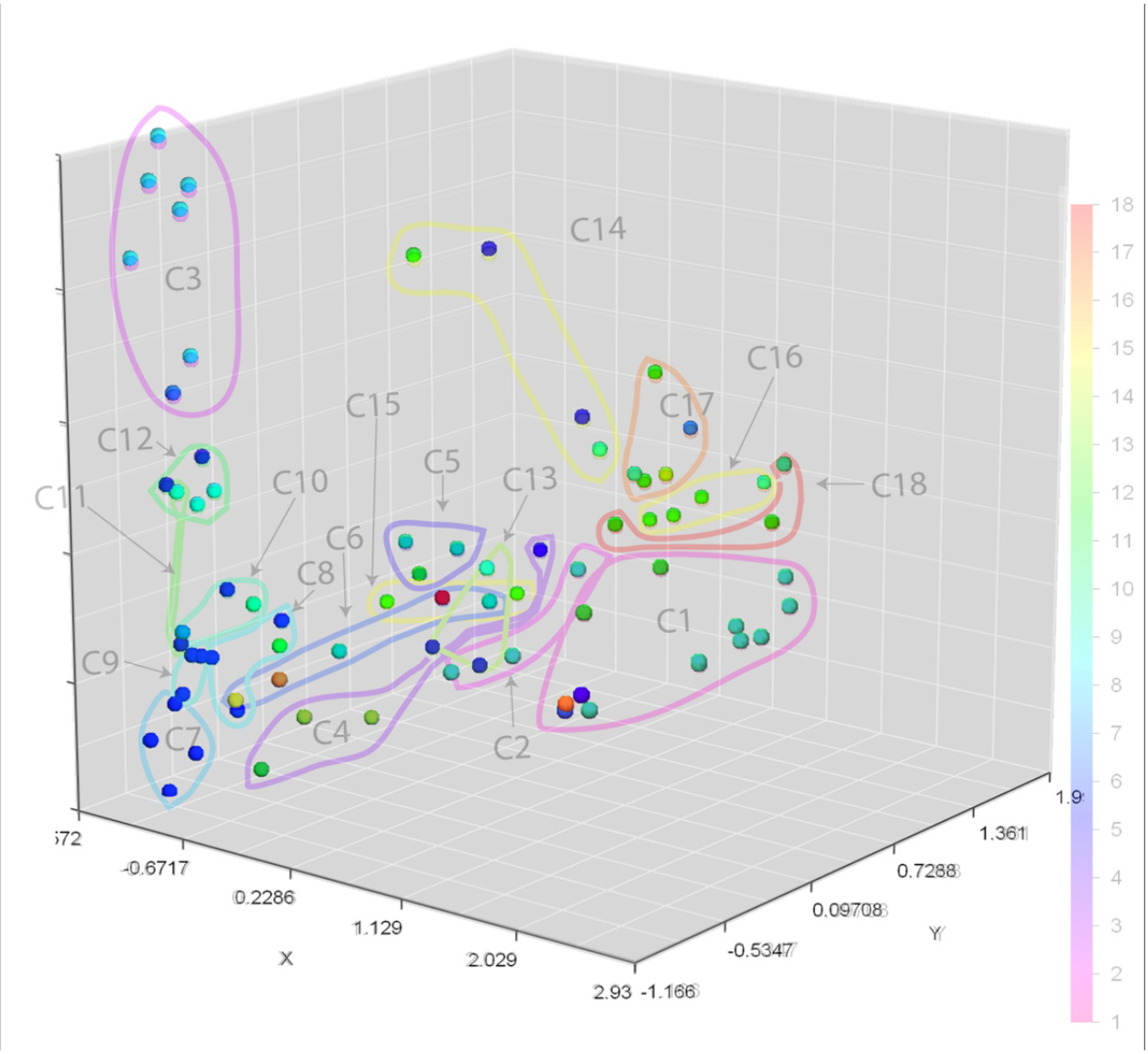

Figure 5 Research Field and Cluster Overlaid Model

Discussion and Conclusion

Multidimensional Scaling, used as a methodological tool, provides constructive analysis of the relationships between the most published authors in a given field (Web 2.0). The findings confirm a foundation within the Computer Science and Library and Information Science fields, however also displayed a significant community of authors within the Medicine/Health, Biomedical, and Bioinformatics fields. Unlike the intertwined relationship of the Medicine/Biomedical/ Bioinformatics, Computer Science/Library and Information Science do not appear as tangled. A clustering analysis found 18 subgroupings within a hierarchical framework. Some of the clusters included authors from unaligned fields, displaying the interdisciplinary nature of those authors.

Overall, the case study successfully demonstrates the use of MDS as a methodology. The resulting visualization illuminates unanticipated relationships, and provides unseen information. Furthermore, the inclusion of three-dimensional modeling tools allows better manipulation of the low-dimensional space. While the study highlighted the interdisciplinary nature of Web 2.0 technologies, its relative newness limited a more robust understanding. The sample population required using authors with as few as two publications on Web 2.0, thus limiting the authors' relationship to Web 2.0. Future research on more established subjects would best illustrate the limitation. Additionally, the limitation to only ISI ranked journals, due to the use of Web of Knowledge, may preclude some authors from inclusion in the study. The addition of non-ISI journals,

however, requires extensive additional variables, such as the selecting which journals to include/exclude, thus making the task ineffective.

References

Biglan, A. (1973). The characteristics of subject matter in different academic areas. *Journal of Applied Psychology*, 57 (3), 195-203.

Bowen, W. M., Dunn, R. A., & Kasdan, D. O. (2010). What is "Urban Studies"? Context, internal structure, and content. *Journal of Urban Affairs*, 32 (2), 199-227.

Culnan, M. J. (1986). The intellectual development of management information systems, 1972-1982: A co-citation analysis. *Management Science*, 32 (2), 156-172.

Easley, S. P. (1990). Multidimensional scaling of social relationships in baboons. *International Journal of Anthropology*, 5 (4), 309-324.

Fenton, M. (1988). Multidimensional scaling and tourism research. *Annals of Tourism Research*, 15 (2), 236-254.

Guttman, L. (1968). A general nonmetric technique for finding the smallest coorinate space for a configuration of points. *Psychometrika* , 33, 469-506.

Mainfort Jr., R. C. (2003). An ordination approach to assessing late period phases in the central Mississippi valley. *Southeastern Archaeology*, 22 (2), 176-184.

Pardoe, I. (2004). Multidimensional scaling for selecting small groups in college courses. *The American Statistician*, 58 (4), 317-321.

Torgerson, M. S. (1952). Multidimensional scaling: I. Theory and method. *Psychometrika* , 17 (4), 401-419.

Young, G., & Householder, A. S. (1941). A note on multidimensional psychophysical analysis. *Psychometrika* , 6 (5), 331-333.

Zhang, J. (2008). Multidimensional Scaling. In J. Zhang, *Visualization for Information Retrieval* (pp. 143-163). Berlin: Springer-Verlag.

Zhang, J., Wolfram, D., & Wang, P. (2009). Analysis of query keywords of sports-related queries using visualization and clustering. *Journal of the American Society for Information Science and Technology*, 60 (8), 1550-1571.

Zhang, J., Wolfram, D., Wang, P., Hong, Y., & Gillis, R. (2008). Visualization of health-subject analysis based on query term co-occurrences. *Journal of the American Society for Information Science and Technology*, 59 (12), 1933-1947.