

# Who we are: Analysis of 10 years of the ED-MEDIA Conference

Xavier Ochoa, Gonzalo Méndez  
Escuela Superior Politécnica del Litoral, Guayaquil, Ecuador  
{xavier, gmendez}@cti.espol.edu.ec

Erik Duval  
Dept. Computerwetenschappen, Katholieke Universiteit Leuven, Leuven, Belgium  
erik.duval@cs.kuleuven.be

**Abstract:** The field of educational multimedia, hypermedia and communications has been explored for more than 15 years in the ED-MEDIA conference. This paper presents a bibliometric and scientometric analysis of the researchers and papers that have been part of ED-MEDIA during these years. The authorship, co-authorship, citation and co-citation analysis are conducted to evaluate the characteristics of ED-Media and its community. The findings suggest that ED-Media is a well balanced scientific venue with a vibrant and collaborative community around it. The metrics obtained are also used to propose a Personalized Recommender System to help the attendants get the most out of the conference.

## 1. Introduction

Ed-Media started as an international conference on ‘Educational Multimedia, Hypermedia & Telecommunications’ in 1993, as a follow-up to an earlier series of International Conferences on Computers and Learning (ICCAL), which started in 1987. The idea was to create a multi-disciplinary forum for the discussion of research and development in this area. It is certainly one of the larger and more international conferences in the field with regularly around 1500 participants from some 70 different countries.

Like any scientific venue, ED-Media can be described through a series of characteristics like the number of published papers, the number of authors, the distribution of papers per author or community, co-citation networks, etc. These characteristics are studied by the fields of Bibliometrics and Scientometrics (Osareh, 1996, Osareh, 1996a, Hood & Wilson, 2001). Bibliometrics measures the production and consumption of scientific material, while Scientometrics is more interested in the measurement of the scientific process. Together, these two sub-fields of Informetrics have developed the methodologies and tools needed to gain insight in conferences, journals and research groups. This paper presents a bibliometric and scientometric analysis of the last 10 years of the ED-Media conference.

The structure of the paper is as follows: section 2 discusses how the data was captured and processed. Section 3 presents the methodology and results of four types of analysis: authorship, co-authorship, citation and co-citation. Section 4 presents a Personalized Recommender System for ED-Media, based on our analysis and metrics. The paper closes with conclusions and opportunities for further work.

## 2. Data Collection and Processing

The main source for the data used in this paper was the Digital Library of the AACE [] and the Proceedings CDs of the ED-Media Conference. The first source, the Digital Library, was used to extract the information about the title, authors and affiliation for each paper from ED-Media 1999 to 2008. The second source, the Proceedings CD was used to extract the information on bibliographical references for each paper from ED-Media 2005 to 2008.

The extraction of the data from the Digital Library was made by web-scraping from the ED-Media pages (AACE, 2008). Small scripts in Java were used to convert those pages from HTML to structured XML documents. For each paper, a list of the authors, affiliation, country and year was extracted. This information is mainly used in the authorship and co-authorship analyses.

The main problem that we faced while extracting this information was that the variability of the affiliation of the authors is presented in substantially varying ways. In some cases, it included the Department, Faculty,

University, city and country. In others there is just the country. Also, the way in which the names were formatted varied a lot between ED-Media editions. For example, initials for the middle name are sometimes included and sometimes not. Most of these problems were solved through the use of text similarity algorithms, like edit distance (Ristad & Yianilos, 1998), or by the removal of obvious errors with a list of reserved words (For example, the words “University” or “Department” in the field author name). In a manual evaluation of 100 papers, the percentage of errors was below 5%.

On the other hand, the extraction of the bibliographical references for each paper from the Proceedings CD was considerably more complex. First, the content table was obtained from the START.PDF file present in the proceedings. This table helped us to divide the main proceedings file, PROCBOOK.PDF, in individual papers. These individual papers were converted to text files using an open source tool called pdftotext, which is part of a larger package called xpdf<sup>1</sup>. For the reference extraction, two main tools were evaluated. The first was ParaTools v1.10<sup>2</sup>, a set of Perl modules used in the ParaCite site<sup>3</sup>. The second tool was parsCit<sup>4</sup>, another set of Perl modules developed by the University of Singapore. Both toolsets were tested over the ED-Media data, but both had advantages and shortcomings. Because most errors were complementary (usually one tool had the reference right when the other failed) we decide to combine the output of both tools to produce the final results. A manual evaluation of 100 papers showed that the percentage of correctly extracted references was around 70%. Circa 30% of the references were not extracted by any tool. The main reason for this failure rate is the non-standard way in which some ED-Media papers cite their references.

Finally, 6,690 papers and 10,689 authors from 92 different countries were identified for the period between 1999 and 2008. Also, 2,946 papers were analyzed from the period 2005-2008 to extract a total of 35,347 citations to 26,378 different papers. This data is used in the following analysis to gain insight into the characteristics of ED-Media.

### 3. Analyses

#### 3.1. Authorship Analysis

Over the last 10 years, ED-Media had an average of 669 papers per conference. Table 1 presents the actual number of papers for each edition of ED-Media and the location where it was held. An interesting trend is that conferences taking place in Europe usually publish a higher number of papers than their North American counterparts. Why this is the case, is not immediately clear to us.

Year	Location	Number of Papers
1999	Seattle, USA	552
2000	Montreal, Canada	603
2001	Tampere, Finland	628
2002	Denver, USA	533
2003	Honolulu, USA	803
2004	Lugano, Switzerland	925
2005	Montreal, Canada	806
2006	Orlando, USA	498
2007	Vancouver, Canada	702
2008	Vienna, Austria	940

<sup>1</sup> Xpdf Web site: <http://www.foolabs.com/xpdf/download.html>

<sup>2</sup> ParaTools Web site: <http://paracite.eprints.org/developers/>

<sup>3</sup> ParaCite Web site: <http://paracite.eprints.org/>

<sup>4</sup> ParsCit Web site: <http://aye.comp.nus.edu.sg/parsCit/>

**Table 1. Number of Papers per ED-Media Conference**

We counted the number of papers published by each individual author over the last 10 years. The first useful information obtained was the rank of most prolific authors. Table 2 presents the top 20 authors by number of papers. This table also presents the number of papers that they got accepted each year. An interesting fact that can be obtained from the numbers in this table is that no author in ED-Media contributes more than 1% of the number of papers of any given conference. Moreover, the total output of the most prolific author, Catherine McLoughlin, is less than 0.6% of the total output of the conference in the last 10 years. As a conclusion, it can be said that ED-Media is not dominated at all by any given author or group of authors.

<b>Author</b>	<b>'99</b>	<b>'00</b>	<b>'01</b>	<b>'02</b>	<b>'03</b>	<b>'04</b>	<b>'05</b>	<b>'06</b>	<b>'07</b>	<b>'08</b>	<b>Total</b>
Catherine McLoughlin	2	6	6	7	5	3	2	1	2	2	<b>36</b>
Ron Oliver	2	4	5	5	5	5	3	0	3	0	<b>32</b>
Joe Luca	1	3	6	4	5	3	3	1	2	1	<b>29</b>
Kanji Akahori	0	0	1	2	1	3	7	4	5	6	<b>29</b>
John Hedberg	2	3	4	2	2	3	6	0	1	1	<b>24</b>
Jan Herrington	2	3	2	1	2	5	2	1	3	3	<b>24</b>
Erik Duval	0	0	3	3	3	3	5	2	1	2	<b>22</b>
Yoneo Yano	2	1	1	2	3	6	5	0	0	2	<b>22</b>
Barry Harper	2	2	3	2	1	5	2	1	0	3	<b>21</b>
Philip Barker	1	1	4	4	1	2	2	2	1	1	<b>19</b>
Karen Swan	0	4	1	3	2	0	3	0	3	3	<b>19</b>
Michael Herczeg	1	3	1	2	4	1	1	1	2	2	<b>18</b>
Min Liu	2	1	2	2	2	1	0	5	2	0	<b>17</b>
Hermann Maurer	3	1	5	2	0	2	0	1	3	0	<b>17</b>
Marco Ronchetti	0	0	0	1	2	3	5	0	3	3	<b>17</b>
Carmel McNaught	0	3	1	1	2	2	4	1	2	1	<b>17</b>
Lori Lockyer	1	2	1	0	2	7	2	1	0	1	<b>17</b>
Tatsuya Horita	1	1	1	0	3	4	2	1	2	2	<b>17</b>
Mike Keppell	1	3	3	2	2	1	1	2	1	0	<b>16</b>

**Table 2. Top 20 most prolific Authors 1999-2008**

A similar analysis was carried out on the countries from where the different authors contribute. Table 3 presents the number of papers for the top 20 most prolific countries. For this count, if a paper has authors that belong to different countries, the paper is added once to each country that participated. Contrary to the situation for individual authors, one country, USA, concentrates more than a third of the total number of publications. However, this situation is not unique to ED-Media. According to the Essential Science Indicators of the ISI Web of Knowledge from 1997 to 2007 (In-cites.com, 2008), USA produces around 30% of the total number of publications worldwide. The next most productive countries are Japan, Germany and UK with around 6-8% each. While the rankings are not the same, 14 of the ED-Media top 20 countries are also in the ISI Web of Knowledge top 20. In that respect, it can be concluded that ED-Media does not seriously over-represent any country.

<b>Country</b>	<b>Papers</b>	<b>Percentage</b>	<b>Rank EdMedia</b>	<b>Rank ISI</b>
----------------	---------------	-------------------	---------------------	-----------------

USA	2441	36.5	1	1
Canada	542	8.1	2	7
Australia	540	8.1	3	11
Japan	440	6.6	4	2
UK	396	5.9	5	4
Taiwan	351	5.2	6	18
Germany	340	5.1	7	3
Austria	164	2.5	8	>20
Finland	143	2.1	9	>20
Italy	139	2.1	10	8
Greece	132	2.0	11	>20
New Zealand	107	1.6	12	>20
Spain	105	1.6	13	10
South Africa	103	1.5	14	>20
France	90	1.3	15	5
Switzerland	89	1.3	16	10
Netherlands	85	1.3	17	12
Hong Kong	75	1.1	18	6
Belgium	69	1.0	19	20
Brazil	68	1.0	20	>20

**Table 3. Top 20 most prolific countries from 1999 to 2008**

The following analysis was to study the distribution of publications over individual authors. A widely reported phenomenon is that the number of papers per author follows a Lotka Distribution (Yablonsk, 1980), also known as Inverse Power Law or Pareto Distribution. If the percentage of ED-Media authors that has produced  $x$  papers is plotted against  $x$  in log-log scales, we find the typical straight line for a Lotka Distribution. Alpha, the unique parameter of this distribution, roughly similar to the slope in the log-log graph, is estimated by the MLE (Goldstein et al., 2004) approach as 2.59. This value falls neatly between the 2 to 3 range found in the majority of scientific publication venues. This distribution also allows us to measure participation inequality: 75% of the authors have contributed only 1 paper, 14% have contributed 2 papers, 5% have contributed 3 papers and just 6% have contributed more than 4 papers. From these percentages, it is easy to determine that the influx of newcomers is an important feature of ED-Media. However, if the cumulative distribution is plotted (Figure 2 right), the concentration, typical also for long tailed distribution can be easily observed. The most productive 20% of the authors are responsible for almost half of the papers published in ED-Media. On the tail however, the least productive 75% (the authors that contribute only 1 paper) are responsible for almost the complete other half. This balance between regulars and new contributors is almost 50-50.

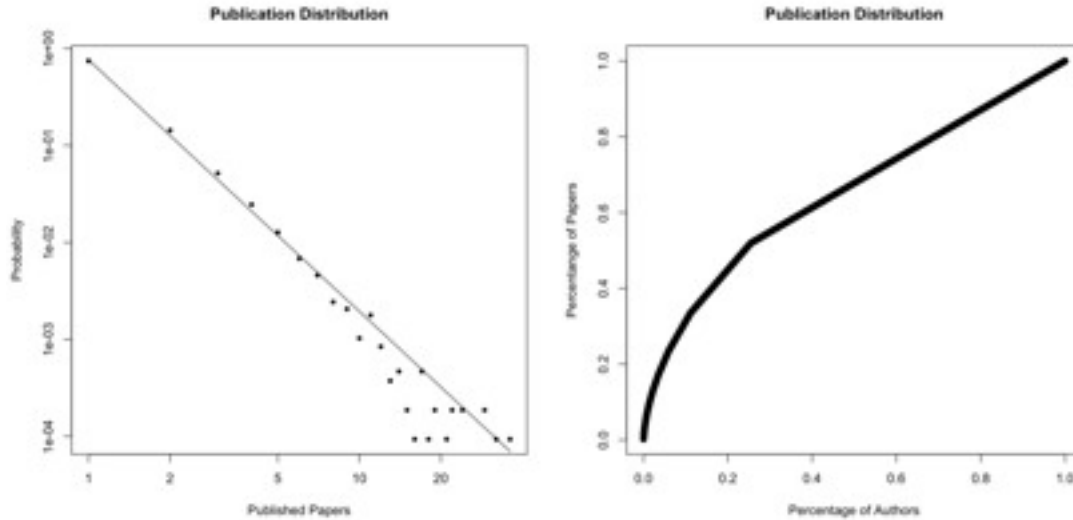


Figure 1. Distribution of Papers per Author. Line is Lotka Distribution with  $\alpha$  2.59

### 3.2. Co-authorship Analysis

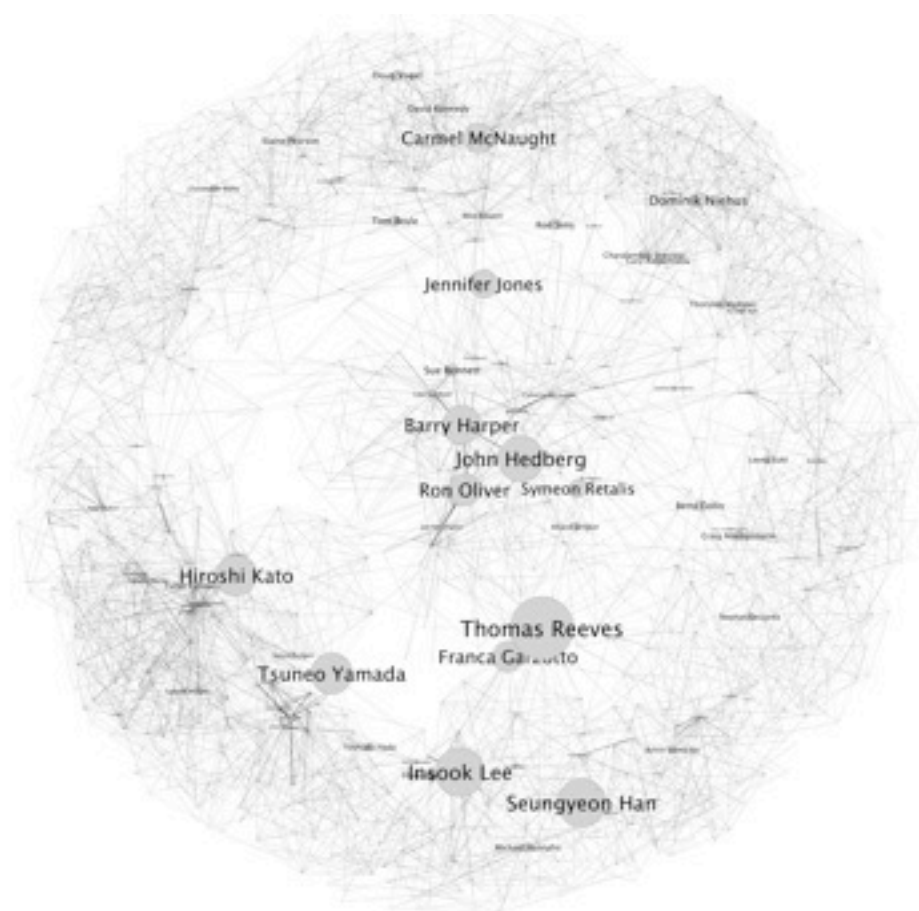
The next set of analyses was focused on the relation between authors and countries as reflected by the papers that have been created in collaboration. For these analyses, all the papers between 1999 and 2008 were considered. From the set of 6690 papers, 2266 (34%) were created by just one author and are not considered in this study. The detail of the number of authors per paper can be seen in Figure 2 (note the log scale in the Y axis). The number of authors per paper decreases exponentially. It is rare (less than 5% of the total) to find papers with more than 6 authors.



Figure 2. Number of authors per paper 1999-2008

Given that co-authorship creates a relation between two authors, that relationship was represented in a graph to gain a better understanding of the network of co-authorships in ED-Media. From the analysis of this graph, we found that there are 1134 connected networks of at least 3 authors. These connected networks represent research groups or at least connected researchers. The largest of these connected networks has 1096 members. This large network can be seen as the core community of collaboration in ED-Media. An illustration of this network can be seen in Figure 3. The size distance between the nodes is determined by co-authorship relationship (close nodes have published together or have published with the same researchers). The size of the node and the names is related to the Betweenness Centrality (Freeman, 1977), a measurement of how important the node is for the connection of the network. If these nodes are eliminated, a part of the

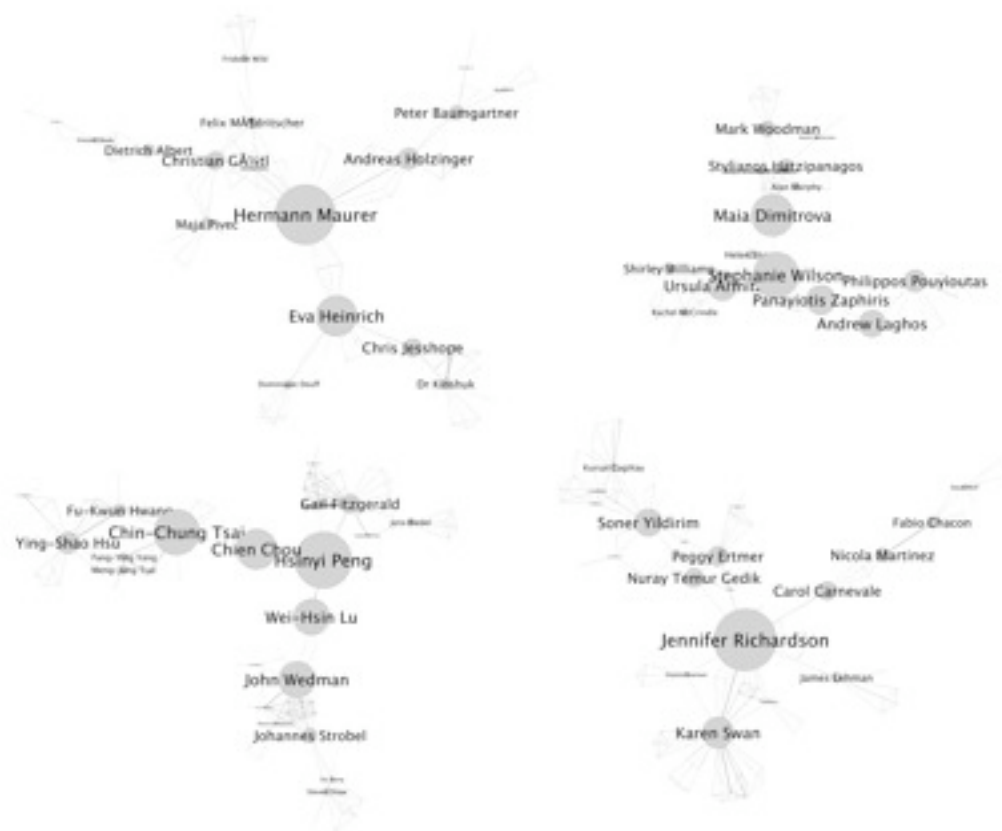
network becomes disconnected from the rest. Note that these authors are not necessarily the most prolific. However, these authors tend to be well “networked” and collaborate with several research groups.



**Figure 3. Biggest Connected Network in the Co-Authorship Graph**

Besides this large network or community, there are smaller connected groups that represent smaller communities or research groups. For example, Figure 4 shows the co-authorship around Hermann Maurer (Austria), Chin-Chung Tsai (Taiwan), Jennifer Richardson (USA) and Maia Dimitrova (UK).

The presence of these groups of different size gives us some insight about the structure of the ED-Media collaboration network. There is a central community where most of the ED-Media “regulars” are immersed. This community is very connected and dynamic with researchers breaking the boundaries of university or institutional groups. There is a second layer of networks that are dominated by one or two researchers, usually well known in ED-Media, that create around them what closely resembles their geographical or institutional networks. There is a third layer of very small groups that represent authors that always publish together without stretching the boundaries of their group.

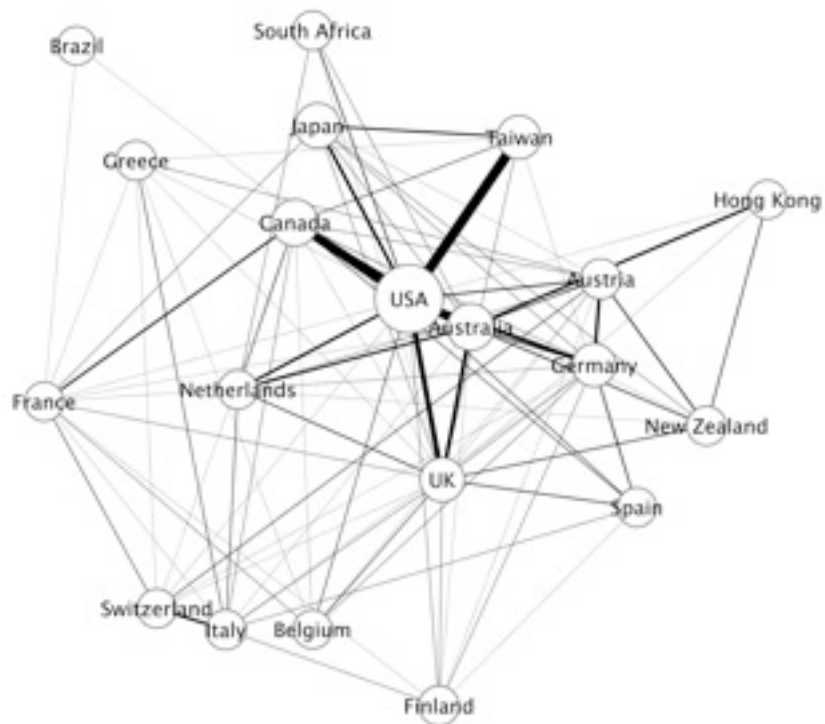


**Figure 4. Other Important Co-authorship Networks**

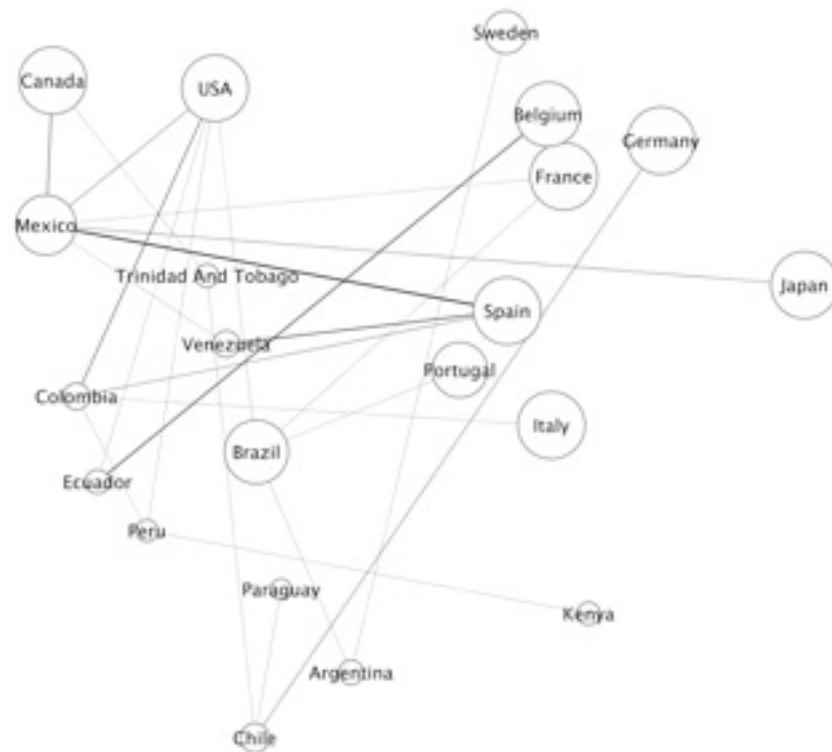
Similar to what we did for the relations between authors, we analyzed the collaboration between countries. Only 580 (9%) papers are the fruit of collaboration between researchers from different countries. The co-authorship graph is one connected network with all the countries that have at least 1 paper in collaboration with another country. The largest collaboration links emanate from USA to Taiwan (30 papers), Canada (27), Australia (24), UK (15) and Germany (11). The largest links that do not include USA are Australia and UK (11 papers), Germany and Austria (9 papers) and Australia and Hong Kong (8 papers). Figure 5 shows the links between the 20 most prolific countries. This result shows that ED-Media researchers publish most often with co-authors from their own countries. Also, when collaboration takes place, it is especially strong between the largest publisher countries.

International collaboration between world regions outside the North America, Europe and Developed Asia and Oceania is rarer. For example, if we analyze the co-authorship network for Latin American countries (Figure 6), it is easy to see that the main collaboration links are between a Latin American country and a developed country (Venezuela and Spain, Ecuador and Belgium, México and USA) but it is almost non-existent among Latin American countries. Maybe ED-Media should encourage special sessions to integrate the representatives from less developed regions of the world in order to foster the creation of collaboration links?

An additional finding is that geography is not a determinant factor when a research partner is selected. Telecommunications have created a truly global research community.



**Figure 5. Collaborative Research links between countries (Top 20 most prolific countries)**



**Figure 6. Latin American Collaboration Network**



### 3.3. Citation Analysis

The next step was to obtain the references from each paper for the last four years (2005 to 2008). These references are usually citations to other papers that support the ideas of the author(s). The first information that can be extracted from this data is the list of the papers most often cited in ED-Media. Table 4 presents the top 15 cited papers or books according to ED-Media 2005 – 2008. This list can be used as a recommended reading list for anyone new to the field of Educational Multimedia.

Author(s), Year	Title	Citations
L. Vygotsky, 1978	Mind in Society: The Development of Higher Psychological Processes	107
E. Wenger, J. Lave, 1991	Situated Learning: Legitimate Peripheral Participation	46
M. Prensky, 2001	Digital Game-Based Learning	43
R. E. Mayer, 2001	Multimedia Learning	37
E. Wenger, 1998	Communities of Practice: Learning, Meaning and Identity	36
A. Collins, J. S. Brown, P. Duguid, 1989	Situated cognition and the culture of learning	34
J. P. Gee, 2003	What video games have to teach us about learning and literacy	28
S. Papert, 1980	Mindstorms: Children, computers, and powerful ideas	27
E. M. Rogers, 1995	Diffusion of Innovations	26
D. Laurillard, 2002	Rethinking university teaching	26
D. H. Jonassen, 2000	Computers as mindtools for schools	24
B. S. Bloom, 1956	Taxonomy of educational objectives: The classification of educational goals: Handbook	19
D. R. Garrison, L. Rourke, T. Anderson, W. Archer, 2001	Assessing teaching presence in a computer conferencing context	19
M. R. Lepper, T. W. Malone, 1987	Making learning fun: A taxonomy of intrinsic motivations for learning	18

**Table 4. Top 15 cited papers or books from EDMedia 2005 – 2008**

Plotting the distribution of citations among referenced papers (Fig. 7- left), we found that it also follows an inverse power law with an alpha of 2.95. This again is expected as scientific citations in other venues [] also follow this distribution with the alpha parameter between 1.5 and 3.5. This is an indication that the citation behavior of ED-Media is similar to most other conferences or journals. The high value of alpha means that the distribution of citations is biased to papers that are cited only once. From the accumulative distribution (Fig. 7 – right) it can be easily seen that 20% of the most cited papers collect almost 40% of the citations. The tail of the distribution, that is papers with just 1 citation, represents the remaining 60% of the citations.

### 3.4. Co-citation Analysis

Two papers can be considered related if they both cite the same paper. The higher the number of cited papers in common, the stronger the relation is. This relation between papers can be transferred also to their authors. In this way, it is straightforward to construct a graph that relates ED-Media authors. Using the information extracted from the papers from ED-Media 2005 to 2008, we created this graph. We found that most authors from 2005 to 2008 (4091) were connected in one giant component, that is, a connected network that contains most of the nodes in a graph. This means that the majority of ED-Media authors are related and that the authors can be considered a coherent community, just by considering the papers that they reference. Moreover, the average distance between any two authors in this connected graph was 4.5. This result shows that the ED-Media community is a sort of small world (Watts & Strogatz, 1998) network with a degree of

separation lower than 5. This author co-citation analysis could be used to find other ED-Media authors that can be working in the same or a similar field. A selection of this graph centered on any of the authors can be extracted. For example, Figure 9 presents the immediate co-citation network of Craig Montgomerie, well known friend of ED-Media. While the most related nodes are Craig's co-authors, there must be researchers there that Craig would want to meet in the next ED-Media, because they are working in a similar field. This idea is further developed in the next section.

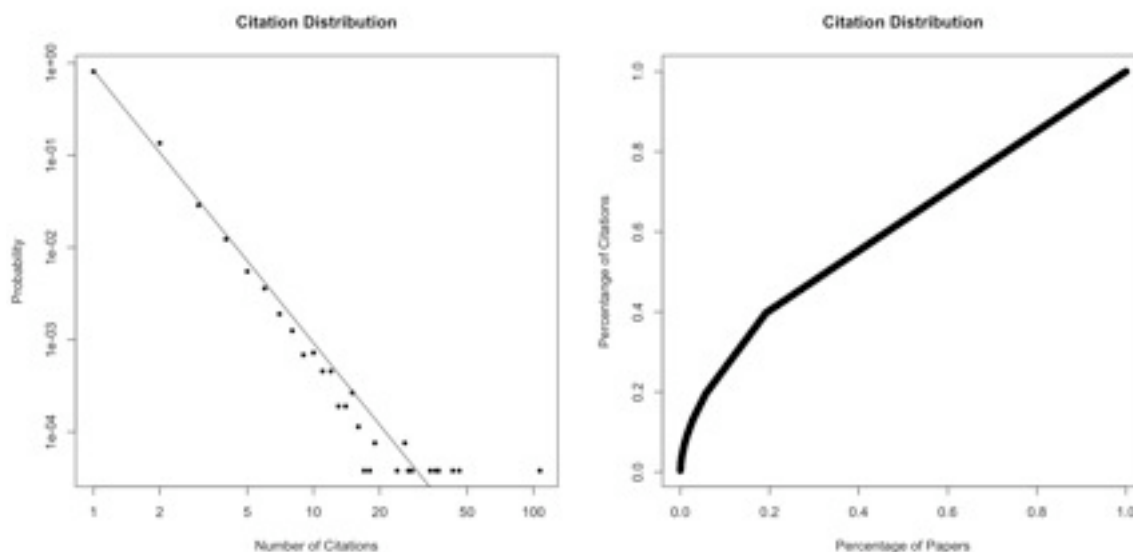


Figure 7. Citation Distribution among referenced papers 2005 - 2008

#### 4. Use of the Results as a Personalized Recommendation System

While knowing about these metrics for the ED-Media conference is interesting, they also have a very practical application to help the researchers that participate at ED-Media, in the form of a Personalized Recommendation System. This system could use information extracted from the previous and current editions of ED-Media to recommend each author about relevant papers, researchers and sessions.

To recommend *papers* to each author, the system extracts the list of the most cited papers in ED-Media, giving extra weight to the papers that have been cited by the co-authors and researchers in the first level of the co-citation network of the author. The system will automatically eliminate from this list the papers that the author has already cited.

To recommend *researchers*, the co-authorship and co-citation graphs are combined. Researchers that have published with the co-authors of the individual and that are not his or her co-authors should be recommended. Also, researchers that have several co-citations in common with the individual and are not in his or her co-authorship network are also good candidates for recommendation.

Once the presentation schedule is ready for the next conference, the system could recommend *sessions* based on the researcher recommendation and also on the co-citation analysis of the paper being presented and the history of ED-Media publications of the attendant.

Based on the presented analyses, the construction of this system is straightforward. This system is being constructed and is expected to be fully functional for ED-Media 2010. For a current demo, that allows ED-Media authors to explore their networks visit <http://nyx.cti.espol.edu.ec/edmedia>.

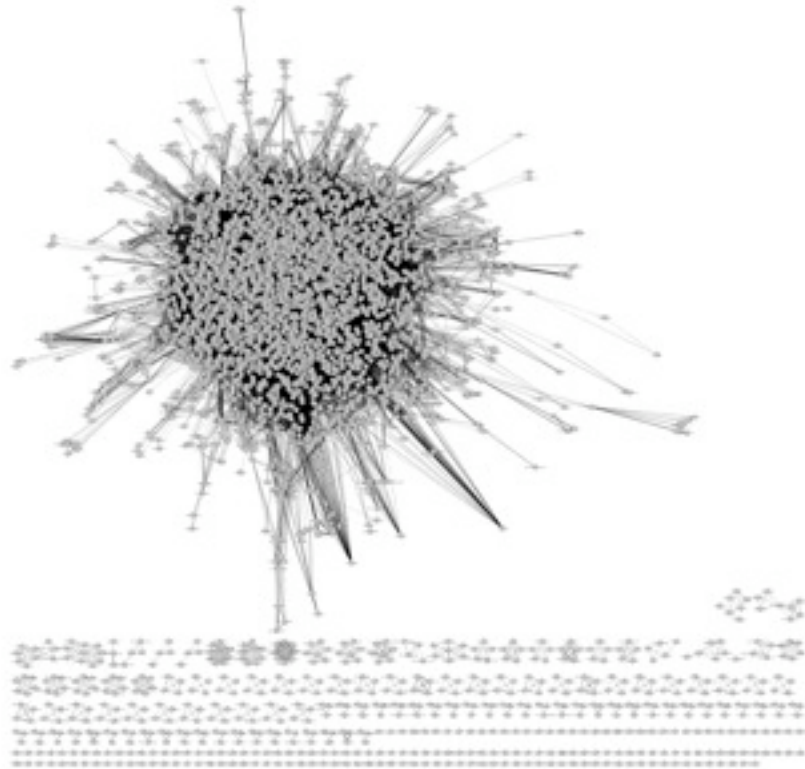


Figure 8. Graph of authors related by Co-citation

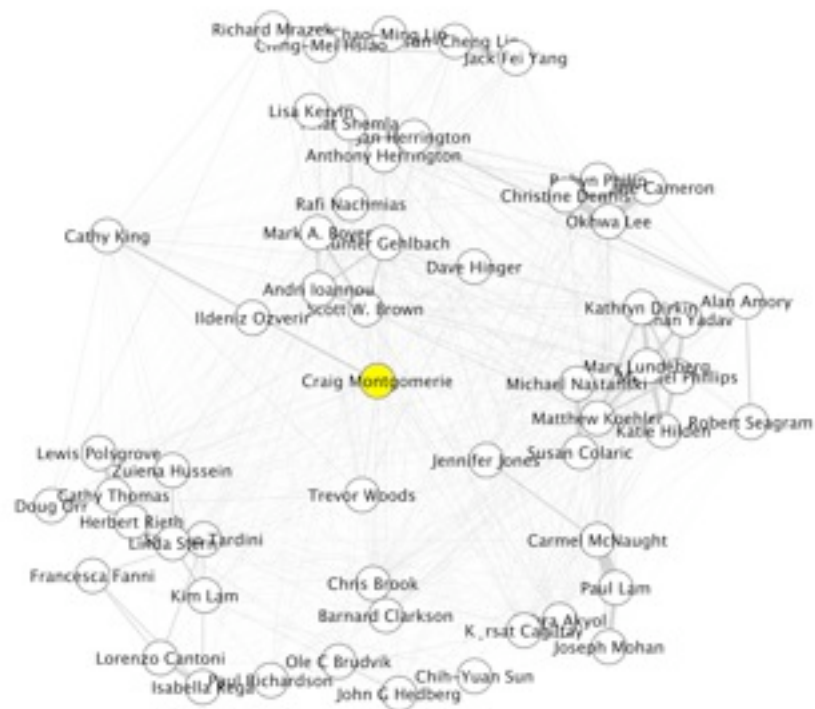


Figure 9. Co-citation network of Craig Montgomerie

## 5. Conclusions and Further Work

The main conclusion from our analysis is that ED-Media is a conference that has the hallmark characteristics of a scientific venue:

- The distribution of publications corresponds to that found in most scientific publications.
- There is a large and dynamic community of collaborating researchers at its core and sub communities defined by institutional and geographical boundaries.
- Its nature is truly international and the country representation follows trends similar to those of other scientific venues.
- The distribution of citations in Ed-Media also has the same distribution and parameters as other citation sources
- Most of the ED-Media authors are related through co-citation, meaning that they form a coherent community.

The information obtained in this analysis can be used to build a Personalized Recommender System for its participants, improving the way that they interact with the conference.

The findings also raise new questions that should be analyzed in further work:

- How does ED-Media compare to other conferences in the same area and other domains?
- How has the ED-Media community evolved over time?
- What is the number and distribution of citations of ED-Media papers in other venues?

Solving these and new questions will help us “get better at getting better” in the research that is conducted in the area of Educational Multimedia and Technology Enhanced Learning in general.

## References

- AACE (2008). ED-Media Conference On-line Library. [Online] <http://www.editlib.org/j/EDMEDIA>
- Coile, R. (1977). Lotka's frequency distribution of scientific productivity. *JASIST*, 28(6)
- Freeman, L. (1977). A set of measures of centrality based on betweenness. *Sociometry*, 40(1), 35-41
- Goldstein, M.; Morris, S. & Yen, G. (2004). Problems with fitting to the power-law distribution. *The European Physical Journal B-Condensed Matter*, 2004, 41(2), 255-258
- Hood, W. & Wilson, C. (2001). The literature of bibliometrics, scientometrics, and informetrics *Scientometrics*, 52(2), 291-314
- In-cites.com. (2007). The year 2007: top list country rankings in all fields. [Online] <http://in-cites.com/countries/2007allfields.html>
- Osareh, F. (1996). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature I. *Libri*, 46, 149-158
- Osareh, F. (1996a). Bibliometrics, Citation Analysis and Co-Citation Analysis: A Review of Literature II. *Libri*, 46, 217-225
- Ristad, E. & Yianilos, P. (1998). Learning string-edit distance. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20 (5), 522-532
- Watts, D. & Strogatz, S. (1998). Small World. *Nature*, 393, 440-442
- Yablonsky, A. (1980). On fundamental regularities of the distribution of scientific productivity. *Scientometrics*, 2(1), 3-34