Journal of Informetrics 8 (2014) 71–76

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Equal weights coauthorship sharing and the Shapley value are equivalent

Alexander Karpov*

National Research University Higher School of Economics, 20 Myasnitskaya Ulitsa, Moscow 101000, Russia

ARTICLE INFO

Article history:
Received 16 April 2013
Received in revised form
21 September 2013
Accepted 25 October 2013
Available online 21 November 2013

Keywords:
Coauthorship
Collaboration
Shapley value

ABSTRACT

The publication credit allocation problem is one of the fundamental problems in bibliometrics. There are two solutions which do not use any additional information: equal weights measure and the Shapley value. The paper justifies the equal weights measure by showing equivalence with the Shapley value approach for sharing co-authors performance in specific games.

© 2013 Elsevier Ltd. All rights reserved.

1. Introduction

Increased collaboration and an increasing number of papers with many co-authors pose a challenge to bibliometrics. In many papers concerning a ranking between authors, research groups, universities, or countries researchers are faced with the problem of publication credit allocation (Egghe, Rousseau, & Van Hooydonk, 2000). Frequently, there is no information about true co-author input. There are several solutions to this problem that do not use any additional information. The first approach is the assignment of full credit to each co-author (normal, or standard counting; Egghe et al., 2000). Lindsey (1980) surveyed 40 papers in social studies of science and the vast majority of them used normal counting. The h-index is based on this assumption (Hirsch, 2005). The second common approach is the equal weights solution (adjusted count; Lindsey, 1980); egalitarian weights; Tol, 2011). The publication credit is divided equally between co-authors. A third approach appeared in recent scientometrics studies (Narayanam & Narahari, 2011; Papapetrou, Gionis, & Mannila, 2011; Sie, Drachsler, Bitter-Rijpkema, & Sloep, 2012; Tol, 2012). Here, the Shapley value solution (Shapley, 1953) was proposed to divide publication credit.

The Shapley value solution originated from cooperative game theory. It was invented to divide benefits from cooperation. For example, person A earns \$100, person B earns \$60. Cooperatively they earn \$180. There are four possible coalitions $\varnothing$, $\{A\}$, $\{B\}$, $\{A,B\}$. Person A could join coalition $\varnothing$ then her marginal contribution would be equal to \$100 or she could join coalition $\{B\}$ then her marginal contribution would be equal to \$180 – \$60 = \$120. Person B could join coalition $\varnothing$ then her marginal contribution would be equal to \$60 or she could join coalition {A} then her marginal contribution would be equal to \$180 – \$100 = \$80. Both agents have equal rights on the benefits from cooperation. This idea is implemented in the Shapley value by calculating the average marginal contribution to all possible coalitions. By the Shapley value, person A in cooperative interaction would have (\$120 + \$100)/2 = \$110. Similarly, person B would have (\$60 + \$80)/2 = \$70.

* Tel.: +7 495 7729590; fax: +7 495 621 1342.
E-mail addresses: akarpov@hse.ru, karpov.hse@gmail.com

1751-1577/\$ – see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.joi.2013.10.008

Table 1
Co-authors' sets.

1	2	3	4	5	6
$\{a\}$	$\{a\}$	$\{b\}$	$\{a,b\}$	$\{a,c\}$	$\{a,b,c\}$

The Shapley value has transparent interpretation in scientometrics studies. Benefit from cooperation is a set of publications written by a set of authors. Authors are players who share common publication credit. Each coalition (subset of authors) corresponds to a number of publication. The Shapley value is equal to the average marginal author's contribution over all possible coalitions. Each author obtains a fair part of the set of publications.

Due to the difficulty of computing the Shapley value, scientometrics studies use some approximations. Tol (2012) used the Shapley value for ranking academic institutions. Tol calculated the average marginal contribution of each researcher but he considered only contribution to existing academic institutions rather than contribution to any possible coalitions. Papapetrou et al. (2011) consider only collaborations of authors who have actually written a paper together, rather than all possible collaborations. Narayanam and Narahari (2011) compute the Shapley value using a sampling-based approach.

Another approach is to find a measure that is equivalent to the Shapley value but is simpler to compute. This solution has been applied to some problems in supply chain management (Chen & Yin, 2010) and phylogenetic biodiversity measures (Hartmann, 2013), but it has not been applied to the co-authorship sharing problem. This paper shows the equivalence of the Shapley value and equal weight sharing. It creates a possibility to count precise theoretical Shapley value rather than an approximations (Narayanam & Narahari, 2011; Papapetrou et al., 2011; Tol, 2012). The Shapley value and equal weight sharing have many justifications obtained in different fields.

The Shapley value satisfies some desirable mathematical properties, moreover Shapley (1953) proved that four axioms define the solution uniquely. Firstly, the solution should allocate the whole publication credit (efficiency); secondly, it should allocate credits equally to authors with equal marginal contributions (symmetry); thirdly, it should allocate zero credit to author without any publications (dummy); fourthly, it should be additive, i.e. for any two disjoint publication sets the value calculated on the whole set should be equal to the sum of the first set value and the second set value (additivity). There are several other axiomatic justifications for the Shapley value. Hart and Mas-Colell (1989) described the Shapley value as a unique solution satisfying symmetry, efficiency and consistency. Chun (1989) described the Shapley value as a unique solution satisfying efficiency, triviality, coalitional strategic equivalence, and fair ranking. The last of these properties means that relative ranks of the agents belonging to the coalition are not affected by the change in the worth of the coalition and it is quite important for our problem.

Myerson (1977) worked on network-based approach justification. He considered games with cooperation structures described by graphs. In the scientometrics context it could be an co-authorship network. He proved that the Shapley value is a unique fair allocation rule for such problems. Fairness in this context means that that two agents should gain equally from their bilateral cooperation.

There are several reasons to support the equal sharing solution (Lindsey, 1980). An individual's return from a co-authored paper with n authors is approximately $1/n$ times that of a single-authored paper (Sauer, 1988). The proportion of equally credited articles in medical journals has increased during the 2000s (Akhabue & Lautenbach, 2010; Tao, Bo, Wang, Li, & Deng, 2012).

There are other approaches to the co-authorship sharing problem that use additional information. Positionally weighted assignment (Abbas, 2011; Lukovits & Vinkler, 1995; Sekercioglu, 2008), harmonic credit (Hagen, 2013), contribution to an article visibility (Egghe, Guns, & Rousseau, 2013) and giving full credit to the senior co-author (Hirsch, 2010) are a few of these approaches.

This paper is organised as follows. Section 2 shows a clarifying example. A mathematical model including the definition of games and the main theorem is presented in Section 3. Section 4 concludes the paper.

2. Example

This section provides an example illustrating the co-authorship sharing problem. Assume there are three authors $N = \{a, b, c\}$. These authors have produced six publications (Table 1). Publication 1 has one author, but publication 6 has three co-authors. There is no information about the contributions of each of the co-authors to each publication.

The author $\boldsymbol{a}$ is the most distinguished, but the majority of his papers are written in collaboration. The problem is how to share credit for the six publications between the three authors. For obvious reasons, the contribution of the first author is not less than two publications (singly-authored) but not more than five publications (with all publications written singly or in collaboration). Finding the exact number of publications contributed to by each author requires the use of a model.

The first approach taken uses the equal weights measure of performance. Credit for each publication is divided equally among its co-authors. Table 2 presents the obtained publication credit allocation $y^{EW} = (3.33, 1.83, 0.83)$.

The Shapley value-based approach appears different. First of all, we should define a correspondence between each subset of authors including empty set and number of publication. There are two main approaches. The first approach is called full obligation game. Each co-author is essential to publication. The publication credit is granted to coalition if all co-authors

Table 2
Equal weights measure of performance.

Authors	{a}	{a}	{b}	{a,b}	{a,c}	{a,b,c}	y_i^{EW}
a	1	1	0	1/2	1/2	1/3	3.33
b	0	0	1	1/2	0	1/3	1.83
c	0	0	0	0	1/2	1/3	0.83

Table 3
Full obligation game.

S	Ø	{a}	{b}	{c}	{a,b}	{a,c}	{b,c}	{a,b,c}
$v_1(S)$	0	2	1	0	4	3	1	6

Table 4
Full obligation game Shapley value.

Permutations	Author ***a***'s contribution	Author ***b***'s contribution	Author ***c***'s contribution
abc	2	2	2
acb	2	3	1
bac	3	1	2
bca	5	1	0
cab	3	3	0
cba	5	1	0
Shapley value	3.33	1.83	0.83

Table 5
Full credit game.

S	Ø	{a}	{b}	{c}	{a,b}	{a,c}	{b,c}	{a,b,c}
$v_2(S)$	0	5	3	2	6	5	4	6

Table 6
Full credit game Shapley value.

Permutations	Author ***a***'s contribution	Author ***b***'s contribution	Author ***c***'s contribution
abc	5	1	0
acb	5	1	0
bac	3	3	0
bca	2	3	1
cab	3	1	2
cba	2	2	2
Shapley value	3.33	1.83	0.83

of the publication belong to the coalition. The function $v_1(S)$ specified in Table 3 assigns a number of publications to each subset of authors S (coalition) for full obligation game.

By definition, the Shapley value is the average marginal contribution added by author i to the coalition. In Table 4, all possible permutations are considered. The marginal contribution of each author is calculated for each permutation.

Let us explain the calculation of the values in Table 4, e.g., the third line. Author ***b*** is the first in the permutation. She takes marginal contribution from $v_1(\varnothing)$ to $v_1(\{b\})$, which is equal to $v_1(\{b\})=1$. Author ***a*** is the second in the permutation, and she receives $v_1(\{a,b\})-v_1(\{a,b\})=3$. Author ***c*** receives residual 2 publications. Taking into account all possible permutations.

The full credit game is based on the maximal number of publications associated with each coalition. The presence of only one author from the set of co-authors is sufficient for publication credit. The function $v_2(S)$ is given in Table 5.

Performing the same calculations as in Table 4, the Shapley value is obtained in Table 6. Moreover, Table 6 can be derived by permutation of the lines of Table 4. This is not coincidental. The Shapley values of dual games are always equal (Funaki, 1998). By definition, one can show the duality of the games v_1 and v_2 in this example.

This example shows the equivalence of the equal weights solution and the Shapley value for basic games. The subsequent section provides a general equivalence result.

3. Results

A mathematical model of the co-authorship sharing problem is described by a set of authors $N=\{1, \ldots, n\}$ and a set of publications $P=\{1, \ldots, m\}$. Each paper is associated with a set of co-authors $S_j \in 2^N$ and a real number quality score $q_j \in R$. It is possible to use the number of citations as the quality measure or to equate the quality of all papers to one. The first case is associated with the problem of citation sharing. In the latter case, only the number of publications should be divided between authors. The aim of modelling this problem is to find the performance measure attributed to each author ($y_i \in R$) subject to one constraint. The sum of the authors' performance measures should be equal to the sum of the publications' quality

$$\sum_{i=1}^{n} y_i = \sum_{j=1}^{m} q_j. \tag{1}$$

There are two main solutions to this problem: equal weights and Shapley value.

Equal weights measure of performance. The quality of a paper (q_j) is divided equally between co-authors

$$y_i^{EW} = \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} q_j, \tag{2}$$

where $|S_j|$ is the cardinality of the set S_j, the number of co-authors and $1_{i \in S_j} = 1$ if $i \in S_j$ and $1_{i \in S_j} = 0$ otherwise.

Before defining the Shapley value, a description of appropriate cooperative games should be given. A cooperative game is described by a characteristic function $v(S)$ that assigns some real number to each coalition. There are three reasonable approaches to assigning this number.

Full obligation game. The absence of one co-author leads to the absence of publication. The game is defined by a characteristic function $v_1(S)$ for all $S \in 2^N$

$$v_1(S) = \sum_{j=1}^{m} 1_{S_j \subseteq S} q_j. \tag{3}$$

Full credit game. Each co-author takes full credit for publication.

$$v_2(S) = \sum_{j=1}^{m} 1_{S_j \cap S \neq \varnothing} q_j. \tag{4}$$

Equal weights game. Each co-author takes an equal portion of the publication's credit.

$$v_3(S) = \sum_{i \in S} \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} q_j. \tag{5}$$

The Shapley value is the average marginal contribution added to the coalition by author *i*. One possible definition of the Shapley value is the average marginal contribution obtained by a randomly arriving player. Let σ be the permutation of players, and let $\sigma(i)$ be the placement of player *i* in permutation σ. The marginal contribution of player *i* is $v(S^{\sigma}_{\sigma(i)}) - v(S^{\sigma}_{\sigma(i)-1})$, where $S^{\sigma}_{\sigma(i)-1}$ is the set of the first $\sigma(i)-1$ players in permutation σ (the set of players in *N* who precede player *i* in the order σ). Taking into account the number of possible permutations, the formula for calculating the Shapley value is derived as follows:

$$\phi_i(v) = \frac{1}{n!} \sum_{\sigma} [v(S^{\sigma}_{\sigma(i)}) - v(S^{\sigma}_{\sigma(i)-1})]. \tag{6}$$

Thus, there are four different methods for measuring an author's performance but they all lead to the same result.

Equivalence theorem. For all co-authorship sharing problems, $\phi_i(v_1) = \phi_i(v_2) = \phi_i(v_3) = y_i^{EW}$.

Proof

Taking into account the definitions of games given above, we rewrite the Shapley values:

$$\phi_i(v_1) = \frac{1}{n!} \sum_{\sigma} \left[\sum_{j=1}^{m} 1_{S_j \subseteq S^{\sigma}_{\sigma(i)}} q_j - \sum_{j=1}^{m} 1_{S_j \subseteq S^{\sigma}_{\sigma(i)-1}} q_j \right] = \frac{1}{n!} \sum_{\sigma} \left[\sum_{j=1}^{m} (1_{S_j \subseteq S^{\sigma}_{\sigma(i)}} - 1_{S_j \subseteq S^{\sigma}_{\sigma(i)-1}}) q_j \right];$$

$$\phi_i(v_2) = \frac{1}{n!} \sum_{\sigma} \left[\sum_{j=1}^{m} 1_{S_j \cap S^{\sigma}_{\sigma(i)} \neq \varnothing} q_j - \sum_{j=1}^{m} 1_{S_j \cap S^{\sigma}_{\sigma(i)-1} \neq \varnothing} q_j \right] = \frac{1}{n!} \sum_{\sigma} \sum_{j=1}^{m} 1_{S_j \cap S^{\sigma}_{\sigma(i)} = \{i\}} q_j;$$

$$\phi_i(v_3) = \frac{1}{n!} \sum_{\sigma} \left[\sum_{i \in S^{\sigma}_{\sigma(i)}} \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} q_j - \sum_{i \in S^{\sigma}_{\sigma(i)-1}} \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} q_j \right] = \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} q_j.$$

Then, $1_{S_j \in S^{\sigma}_{\sigma(i)}} - 1_{S_j \in S^{\sigma}_{\sigma(i)-1}} = 1$ is true for $i \in S_j$, and author *i* is the last author among co-authors S_j in ordering σ.

$1_{S_j \cap S^{\sigma}_{\sigma(i)} = \{i\}} = 1$ is true for $i \in S_j$, and author *i* is the first author among co-authors S_j in ordering σ.

If $i \in S_j$, then author i has an equal chance of being first, second or last among co-authors S_j in the ordering σ. The probability of being last is $1_{i \in S_j}/|S_j|$. The same result can be derived by counting all permutations that satisfy each respective condition. The number of such permutations is the number of ways of placing the unordered set of co-authors (except the author i) of paper j in an ordering from N, multiplying by the number of permutations of co-authors of paper j, and the number of permutation of all authors except the co-authors of paper j

$$\phi_i(v_1) = \sum_{j=1}^{m} \frac{(n/|S_j|) \cdot (n - |S_j|)! \cdot (|S_j| - 1)!}{n!} \cdot 1_{i \in S_j} \cdot q_j = \sum_{j=1}^{m} \frac{1_{i \in S_j}}{|S_j|} \cdot q_j.$$

The same result is true for the probability of being first:

$\phi_i(v_1) = \phi_i(v_2) = y_i^{EW} = \sum_{j=1}^{m} \frac{1_{i \in S_j}}{(S_j)} \cdot q_j$.For all games that do not distinguish all authors and all publications by names or by another characteristic, we obtain the equal weights solution as equivalent to the Shapley value solution. The same result can be obtained using Shapley value axiomatics (see Winter, 2002). The Shapley value satisfies the additivity property:

$$\phi_i(v + w) = \phi_i(v) + \phi_i(w) \tag{7}$$

The Shapley value for the co-authorship sharing game corresponding to the entire set of publications is equal to the sum of the Shapley values of the games corresponding to the problem of dividing one publication

$$\phi_i(v) = \sum_{j=1}^{m} \phi_i(v^j), \tag{8}$$

where v^j is the game corresponding to the sharing of credit for publication j: $v^j(S_j) = 1, v^j(N \backslash S_j) = 0$. Publication credit should be divided between co-authors in a manner satisfying the Shapley value axiomatics: the efficiency, symmetry, dummy, and additivity axioms. A unique method of dividing credit in such a manner is equal dividing.

A Shapley value-based approach can be applied to heterogeneous scholars differentiated by certain weights. These weights are used to determine a probability distribution over orders of authors, and the weighted Shapley value is the expected contribution of the authors (Monderer, Samet, & Shapley, 1992). There are several axiomatic justifications for the weighted Shapley value (Kalai & Samet, 1987; Radzik, 2012). The additivity property weighted Shapley value can be rewritten as the sum of one paper value. According to the weighted approach, each author's contribution is equal to the author's respective weight in the publication. Therefore, the weighted Shapley value is the sum of the author's respective weights in each publication.

$$\phi_i(v) = \sum_{j=1}^{m} \frac{1_{i \in S_j} w_i}{\sum_{k \in S_j} w_k} q_j \tag{9}$$

Formula 9 generalises formula 2 in the case of unequal authors' weights. The equivalent to the weighted Shapley value is a weighted measure where credit for each publication is divided according to the authors' weights. This is a simplified manner of using the weighted Shapley value.

4. Conclusion

This paper fills a gap in the literature by showing the equivalence of two existing solutions. The equivalence theorem provides a game-theoretical justification for the equal weights measure of performance, and it provides a simple method for computing the Shapley value by substituting an equal weights measure. An equal weights measure that retains all properties of the Shapley value can be used. The weighted extension of the Shapley value is considered, and a simplified formula for computing the weighted Shapley value is obtained.

Acknowledgements

The author acknowledges the support of the DeCAn Laboratory (HSE). This article is an output of a research project implemented as part of the Basic Research Programme at the National Research University Higher School of Economics (HSE). The author would like to thank Ivan Boldyrev for inspiring an interest in scientometric problems and Dirk Meissner and two anonymous referees for detailed and useful comments.

References

Abbas, A. (2011). Weighted indices for evaluating the quality of research with multiple authorship. *Scientometrics*, *88*(1), 107–131.
Akhabue, E., & Lautenbach, E. (2010). "Equal" contributions and credit: An emerging trend in the characterization of authorship. *Annals of Epidemiology*, *20*(11), 868–871.

Chen, R. R., & Yin, S. (2010). The equivalence of uniform and Shapley value-based cost allocations in a specific game. *Operations Research Letters*, *38*(6), 539–544.

Chun, Y. (1989). A new axiomatization of the Shapley value. *Games and Economic Behavior*, *1*(2), 119–130.

Egghe, L., Guns, R., & Rousseau, R. (2013). Measuring co-authors' contribution to an article's visibility. *Scientometrics*, *95*(1), 55–67.

Egghe, L., Rousseau, R., & Van Hooydonk, G. (2000). Methods for accrediting publications to authors or countries: Consequences for evaluation studies. *Journal of the American Society for Information Science*, *1*(2), 145–157.

Funaki, Y. (1998). *Dual axiomatization of solution of cooperative games*. Mimeo.

Hagen, N. T. (2013). Harmonic coauthor credit: A parsimonious quantification of the byline hierarchy. *Journal of ICS*, *7*(4), 784–791.

Hart, S., & Mas-Colell, A. (1989). Potential, value, and consistency. *Econometrica*, *57*(3), 589–614.

Hartmann, K. (2013). The equivalence of two phylogenetic biodiversity measures: the Shapley value and Fair Proportion index. *Journal of Mathematical Biology*, *67*(5), 1163–1170.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Science of the USA*, *102*, 16569–16572.

Hirsch, J. E. (2010). An index to quantify an individual's scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, *85*(3), 741–754.

Kalai, E., & Samet, D. (1987). On weighted Shapley values. *International Journal of Game Theory*, *16*(3), 205–222.

Lindsey, D. (1980). Production and citation measures in the sociology of science: The problem of multiple authorship. *Social Studies of Science*, *10*(2), 145–162.

Lukovits, I., & Vinkler, P. (1995). Correct credit distribution: A model for sharing credit among coauthors. *Social Indicators Research*, *36*(1), 91–98.

Monderer, D., Samet, D., & Shapley, L. S. (1992). Weighted values and the core. *International Journal of Game Theory*, *21*(1), 27–39.

Myerson, R. B. (1977). Graphs and cooperation in games. *Mathematics of Operations Research*, *2*(3), 225–229.

Narayanam, R., & Narahari, Y. (2011). A Shapley value-based approach to discover influential nodes in social networks. *IEEE Transactions on Automation Science and Engineering*, *8*(1), 130–147.

Papapetrou, P., Gionis, A., & Mannila, H. (2011). A Shapley value approach for influence attribution. *Machine Learning and Knowledge Discovery in Databases Lecture Notes in Computer Science*, *6912*, 549–564.

Radzik, T. (2012). A new look at the role of players' weights in the weighted Shapley value. *European Journal of Operational Research*, *223*(2), 407–416.

Sauer, R. D. (1988). Estimates of the returns to quality and coauthorship in economic academia. *Journal of Political Economy*, *96*(4), 855–866.

Sekercioglu, C. H. (2008). Quantifying coauthor contributions. *Science*, *322*(5900), 371.

Shapley, L. S. (1953). A value for n-person games/in contributions to the theory of games, vol, II, by H. W. Kuhn and A. W. Tucker (Eds.). *Annals of Mathematical Studies*, *28*, 307–317.

Sie, R. L. L., Drachsler, H., Bitter-Rijpkema, M., & Sloep, P. (2012). To whom and why should I connect? Co-author recommendation based on powerful and similar peers. *International Journal of Technology Enhanced Learning*, *4*(1–2), 121–137.

Tao, T., Bo, L., Wang, F., Li, J., & Deng, X. (2012). Equal contributions and credit given to authors in anesthesiology journals during a 10-year period. *Scientometrics*, *91*(3), 1005–1010.

Tol, R. S. J. (2011). Credit where credit's due: Accounting for co-authorship in citation counts. *Scientometrics*, *89*(1), 291–299.

Tol, R. S. J. (2012). Shapley values for assessing research production and impact of schools and scholars. *Scientometrics*, *90*(3), 763–780.

Winter, E. (2002). The Shapley value. In R. J. Aumann, & S. Hart (Eds.), *The handbook of game theory with economic applications* (vol. 3) (pp. 2026–2052). Amsterdam: North-Holland.