

The effect of collaborators on institutions' scientific impact

Ali Gazni^{1,2}* Vincent Larivière^{3,4}* Fereshteh Didegah⁵

Abstract: The effect of collaborators on institutions scientific impact was examined for 81 institutions with different degrees of impact and collaboration. Not only collaborators including both core and peripheral collaborators cite each other more than non-collaborators, but also the first group cites each other faster than the second group even when self-citations were ignored. Although high impact institutions and more collaborative institutions receive more citations from their collaborators, it seems that the number of these citations increases only up to a certain point. In this regard, for example, there is a slight difference between top and middle collaborative institutions; however, only a small fraction of collaborators do not cite back the papers of these two groups of institutions. The benefit of collaboration varies based on the type of collaborators, institutions, papers, citers and the publication year of cited documents. For example, the effect of collaboration decreases as the institutions level of impact increases. Hence, collaborating more does not directly imply obtaining higher impact.

Keywords: scientific collaboration; collaboration strategies; citation patterns; inter-organization collaborations

Introduction

Over the recent years, research collaboration has been encouraged by many policymakers at the regional, national, and international levels, mostly through funding programs (Defazio et al., 2009). Not only the number of entities (i.e. authors, institutions and countries) involved in research collaborations has been increasing over time (Gazni, Sugimoto, & Didegah, 2012), but we also observe that, on average, scientific impact increases with the number of authors, institutions and countries (Jones, Wuchty, & Uzzi, 2008; Larivière, Gingras, Sugimoto, & Tsou, 2015). However, the effect of collaboration on articles' citation rates varies according to the authors (Levitt & Thelwall, 2009), **regions** (Levitt & Thelwall, 2010), institutions (Jones, Wuchty, & Uzzi, 2008), countries (Zhao & Guan, 2011) involved, as well as to the discipline (Didegah & Thelwall, 2013). Given such variations, it is relevant to have a better understanding of the gains—in terms of citations received—to be obtained from such collaborations. For example, what citation increase can be expected, for an institution that typically has a lower impact, by collaborating with a high impact institution? In other words, does more collaboration always result in higher citations, or does the effect of collaboration on scientific impact varies according to the collaborative institutions?

The speed at which citations are obtained can also be considered as an indicator of the level of interest of the research community for a research paper (van Dalen & Henkens, 2005; Bornmann &

¹ . Member of Research Department of ISC, Regional Information Center for Science and Technology, Shiraz, Iran.

² . Vice President in Research Affairs, ISC, Shiraz, Iran.

³ . École de bibliothéconomie et des sciences de l'information, Université de Montréal, Montréal, QC, Canada.

⁴ . Observatoire des Sciences et des Technologies (OST), Centre Interuniversitaire de Recherche sur la Science et la Technologie (CIRST), Université du Québec à Montréal, Montréal, QC, Canada.

⁵ . Research Unit for the Sociology of Education, University of Turku, Turku, Finland.

Daniel, 2010). As most evaluations and rankings are based on citations received by the most recent publications—despite providing an incomplete portrait of institutions’ global scientific impact—such short term citations can be considered as more important than older ones. In this context, providing empirical insights on the proportion of short-term citations coming from collaborators versus non-collaborators could help understand the factors that influence scientific impact, and the benefit of collaboration. Performed at the level of institutions, the study aims to seek answers to the questions below:

1. *Do collaborators cite institutions’ recent publications faster than non-collaborators?*
2. *Do core-collaborators cite institutions’ recent publications faster than peripheral collaborators?*
3. *To what degree do non-collaborators and each group of collaborators cite institution's publications?*

Literature review

Previous research has shown that, as the number of authors, institutions and countries appearing on a paper increases, its number of citations also increases (Jones, Wuchty, & Uzzi, 2008; Larivière, Gingras, Sugimoto, & Tsou, 2014). Many reasons for the highest impact of collaborative research have been suggested, among which that it has a highest epistemic authority (Wray, 2002), or that it integrates the knowledge of many different people which, in turn, reduce the risk of errors (Thagard, 1997). However, other more sociological factors might also affect papers’ citation rates, such as the awareness, by a larger network of scholars, that a paper exists, which can be linked to collaborators' formal (Goldfinch, Dale, & DeRouen, 2003) and informal communication networks (Laudel, 2002).

At author level, Wallace, Larivière, & Gingras (2012) have shown that an increasing share of citations received is comes from collaborators, as well as collaborators of collaborators. These trends vary significantly among disciplines, with a fewer share of such network citations in history compared to astrophysics. Similarly, Ajiferuke, Lu, & Wolfram (2011) provided evidence that collaborators cite each other's papers more than non-collaborators. Along these lines, Gazni and Thelwall (2014) found that collaborators' number of citations to each other and shared references are also increasing over time. However, such trends varied across collaborators: while some collaborators authors share a large proportion of common citations and references with each other, others do not cite and have no common references with a large group of the collaborators, thus suggesting that different factors shape collaboration and citation behavior. Indeed, as many have shown, social tie is not a sufficient condition to trigger citation between authors, and that intellectual ties are a better predictor (White, Wellman, & Nazer, 2004). Martin et al. (2013) analyzed the speed of citations, and found that self-citations are typically the first citations to be received, and that the authors’ collaborators are more likely to cite slightly sooner than non-collaborators. Similar results were also obtained by Glänzel and Thijs (2004). However, at the level of institutions—which can be considered as authors’ larger social network—little is known on the speed at which collaborators cite each other’s’ works. For instance, Yan & Ding (2012) have shown, using 59 journals in library and information science, that, as the number of collaborations between two institutions increase, the chance of citations between them also increase. Similarly, institutions with a similar research profile were more likely to collaborate with each other more (Thijs & Glanzel, 2010). However, not all institutions have the same importance in the scientific network. For instance, a small number of institutions are very central, while many others are at the periphery (Chang & Huang, 2013). Scientific impact has also proved to be a factor influencing collaborative behavior: institutions with the same level of impact have less tendency for collaboration with each other, while collaboration with high impact institutions results in higher impact papers compared to collaboration with lower impact ones (Ahn, Oh, & Lee, 2014). Gazni and Thelwall (2015) revealed that the number of inter-

organizational collaboration of top 100 high-impact institutions have increased faster than less high-impact ones. They also found that although top 100 institutions collaborate increasingly with other institutions than top 100 ones, they cite the papers of the top 100 institutions more than other institutions, so they prefer to cite a narrower set of high-quality work instead of the research that they are written jointly with less high-impact institutions.

On the whole, these previous studies showed that, on average, multi-institutional papers have higher impact (e.g. Andrade, de Los Reyes Lopez, & Martín, 2009; Gazni & Didegah, 2011; Larivière, Gingras, Sugimoto, & Tsou, 2015). However, they do not provide insights on the influence of different groups of collaborators on institutions' scientific impact.

Data and Methods

1. Examined institutions

This study processed citable documents (i.e. articles, reviews, and proceedings papers) indexed in Thomson Reuters' Web of Science (WoS) over the 2009-2012 period. Such time period was the period covered by the CWTS Leiden ranking 2014—the most up-to-date version at the time this research was conducted. The top 750 institutions with the highest number of WoS papers were classified into nine groups based on the Mean Normalized Citation Score and number of their collaborations with other institutions. These institutions were categorized into top, middle, lower groups with respect to their impact, and into top, middle and lower based on their number of collaborations. These groups thus include 1) top-top, 2) top-middle, 3) top-lower, 4) middle-top, 5) middle-middle, 6) middle-lower, 7) lower-top, 8) lower-middle and 9) lower-lower. For example, top-top refers to top-impact and top-collaborative institutions or lower-middle refers to lower-impact and middle-collaborative institutions. For each group, nine institutions were randomly selected—for a total of 81 institutions. Citations received by papers published by the 81 institutions during the 2009-2012 period were processed. **The 81 institutions published 645,923 papers during 2009-2012 and these papers have been received 9,400,491 citations.**

Unifying institution name variations

Institution names can change over time (i.e. split or merge, open or close) and authors might use different names for their affiliation. WoS provides a list of institutions' name variations to identify the different various ways an institutions' name can be written. The WoS uses addresses, geographic regions, sub-organizations, departments, postal codes, advisors and a database of mergers and name changes to assign an affiliation to an institution⁶. Lists of variants are provided for institutions under in the "Organizations - Enhanced List" section in WoS. This page is available when the Organizations-Enhanced field is chosen in a search. Such list was used to identify the different names under which institutions are written.

2. Counting the number of collaborations

The fractional counting method was used for calculating the number of collaborations of each institution, which does not give equal weight to all institutions on a paper. For example, if a paper was authored by 3 authors from institution A and 2 authors from institution B, then the contribution of

⁶ <http://ip-science.thomsonreuters.com/m/pdfs/UnifyingNameVariants.pdf>

institution A to the paper is $3/5$ or 0.6 and that of institution B is $2/5$ or 0.4. For each year from 2009 to 2012, the number of collaborations of each institution was counted separately. **The 81 institutions have collaborated with 4,940 institutions during 2009-2012.**

3. *Core-peripheral collaborators*

For each institution A and publication year Y, collaborative institutions were determined and, for each collaborative institution, the proportion of collaborations from the institutions A's collaborations in that year were calculated. Then the collaborators in each year Y were sorted by descending frequency of number of collaborations. In each year Y, core-collaborators are those which collectively account for 80% of papers authored in collaboration, and peripheral collaborators are those that fall into the remaining 20% of collaborations.

4. *Citations received by papers*

To determine the speed of citations received (research questions 1 and 2), we compiled the proportion of field and year normalized citations received by each institutions that were made the same year as the publication year, for both core and peripheral citing institutions. In order to assess the effect of institutions' self-citations on these trends, we compiled proportion with and without papers on which the cited institution appeared. To answer the third research question, the proportion of field and year normalized citations made by WoS papers published in each examined year (e.g. 2009) to the papers of each 81 examined institutions irrespective of the year of the cited items (1990-2012) was measured.

Field and year normalization of citations was obtained by compiling the average number of citations in each citing subject-fields and each citing publication year (2009-2012) to each cited year (1990-2012). For example, the average number of citations from mathematics' papers published in 2009 to the papers published in 2009 or 2008 irrespective of the cited papers subject-fields. The 38 OECD subject-fields⁷ were used for journals' classification. Each citation received by the papers of 81 examined institutions was divided by the average number of citations with respect to the citing subject-fields, citing year and the cited item year. For instance, papers in basic medicine published in 2011 cite any other published papers in 2008, on average, about 3.9 times, thus each citation in that field and citing and cited year was divided by 3.9.

5. *Core-peripheral citers*

Institutions that cite institution A in year Y were commonly named citers. For each citer, the proportion of citations made to the institution A among all other citations received by that institution was computed. Then the citers in year Y were sorted accordingly based on the number of citations from the most to the least. In year Y, institutions that collectively accounted for 80% of citations are considered as core-citers, while those accounted for 20% of citations are considered as peripheral.

6. *Collaborators and citers classifications*

The collaborators and citers were classified into eight classes as follows based on the number of collaborations and citations:

- a. Core-collaborators and core-citers

⁷. http://incites.isiknowledge.com/common/help/h_field_category_oecd.html

- b. Core-collaborators and peripheral-citers
- c. Peripheral-collaborators and core-citers
- d. Peripheral-collaborators and peripheral-citers
- e. Core-collaborators and non-citers
- f. Peripheral-collaborators and non-citers
- g. Core-citers and non-collaborators
- h. Peripheral-citers and non-collaborators

Results

Do collaborators cite institutions' recent publications faster than non-collaborators?

72% of citations received by institution's recent papers come from its collaborators while only 28% of citations to the institution are from non-collaborators (Figure 1). When the institution's self-citations were excluded (Figure 2), the proportion citations coming from collaborators' decrease to 58%, thus showing that an important share of collaborators' citations to each other works are made to papers they jointly authored. Unsurprisingly, all groups of institutions receive a larger proportion of their citations from their collaborators than non-collaborators. However, collaborators from top and middle collaborative institutions cite them at a much higher rate than collaborators from lower collaborative institutions. On average, 76% of citations received by top collaborative institutions come from their collaborators, while these percentages are of 75% and 64%, respectively, for middle and lower collaborative institutions (Figures 3, 4 and 5). Comparing the results in Figures 3, 4 and 5 with that of Figures 6, 7 and 8 reveals that institutional self-citations are more common in lower collaborative institutions than in the top and middle ones.

The collaborators of high and middle impact institutions cite them much more than the collaborators of lower impact institutions (Figures 3, 4 and 5). For example, on average, about 81% of citations to high impact institutions come from their collaborators, while only 62% of citations to lower impact institutions come from their collaborators. This fact is true even when the institutional self-citations were excluded, so from another perspective, on average, lower impact institutions receive more citations from their non-collaborators compared to the high and middle impact institutions. On average, lower, middle and high impact institutions respectively receive 56%, 41% and 28% of the citations from their non-collaborators (Figures 6, 7 and 8).

Do core-collaborators cite institutions' recent publications faster than peripheral collaborators?

A comparison of citations to the institutions' papers (Figures 3, 4 and 5) and citations to the jointly written papers with the institutions (Figures 6, 7, and 8) shows that peripheral collaborators also cite institutions' papers more than those with which there are no collaborations. This is true for all groups of examined institutions; however, institutions that are more collaborative relatively receive more citations from peripheral collaborators. Moreover, as the institutions' level of impact and collaboration increase, the number of citations received from core-collaborators also increases, except for high-impact / middle-collaborative institutions—which receive the highest proportion of citations from their core-collaborators (Figures 3, 4 and 5). This is also observed even when institution's self-citations are removed (Figures 6, 7, and 8).

A comparison of Figures 6, 7 and 8 also shows that core-collaborators of lower-collaborative / lower-impact institutions do not cite back the jointly written papers with the institution as the core-collaborators of other groups of the institutions do. Middle collaborative institutions also receive more citations from their collaborators than other groups do (Figures 6, 7 and 8). Middle-collaborative / high-impact and top-collaborative / high-impact institutions receive more citations from their core-collaborators than other groups of institutions receive (Figures 3, 4 and 5).

Peripheral-collaborators could be classified into three groups according to their number of citations to the institution: core-citers, peripheral-citers and non-citers. Although on average, about 13% of peripheral-collaborators do not cite the institution's publications—especially for lower-impact institutions (about 15% of peripheral-collaborators)—Figures 3 to 8 show that peripheral-collaborators made a considerable number of citations to the institution and approximately the number of citations increased as the institution's impact increased.

Figure 3 to 8 reveal that, on average, lower-impact institutions gain more benefit from peripheral-citers and also core-citers and non-collaborators compared to the higher-impact ones. These Figures also demonstrate that lower-collaborative institutions receive more citations from core-citers and non-collaborators group than the top-collaborative institutions.

To what degree do non-collaborators and each group of collaborators cite institution's publications?

A comparison of Figure 1 and 9 display that collaborators cite institution's most recent publications more than the institution's publications in any years (71% versus 62%); however, for both groups of publications, the collaborators cite the institution more than non-collaborators (Figure 1 and 9). Investigating core-collaborators that do not cite the institutions' publications reveals that, especially for middle and top collaborative institutions, only very few core-collaborators (about 1%) do not cite the institutions' publications. Those core-collaborators that are also core-citers cite the institutions' publications from any year more than the most recent publications of the institution (Figure 1 and 9). Core-collaborators-non-citers group consists of a smaller proportion of collaborators for high-impact institutions (near the zero percent) than the lower-impact ones (about 2%).

Discussion and conclusion

Results show that institutions receive the highest credit from their collaborators while collaborators extensively cite the papers that they themselves are collaborating in; however non-collaborated papers were cited more than the collaborated ones. Collaborators cite 71% of the institutions' most recent publications (Figure 1), while 58% of them are non-collaborated papers (Figure 2). Collaborations are not always rewarded by co-authorship (Laudel, 2002) and the citers of non-collaborated papers might be informed about their partners' other research works when they are working together informally. Along these lines, collaborators are also quite likely to work on similar lines of research (Thijs & Glanzel, 2010).

Although collaborators cite institutions' papers much more than non-collaborators, being highly collaborative does not lead to receiving a higher number of citations from collaborators. In other words, it seems that the number of citations received from collaborators increases, but only up to a certain point. Figures 3 to 5 shows that, on average, the more collaborative the institution, a higher number of citations it receives from collaborators; however, there is only a slight difference between top and middle

collaborative institutions. Moreover, Figures 6-8 and 10-12 reveal that there is no difference between top and middle collaborative institutions based on the number of citations they received from the collaborators. However, more collaborative institutions greatly benefit from their collaborators as they receive a large number of their citations from their collaborators and that collaborators do not only cite their joint publications with the institution, but are also citing other institution's publications (Compare Figures 3-5 and 6-8).

Figures 1, 2 and 9 show that collaborators have an important role in increasing institution's impact, whereas Figures 3-8 and 10-12 reveal that doing a successful research is more important than being only more collaborative. Nevertheless, as a collaboration strategy, choosing to be top, middle or lower collaborative might affect the impact of the institution, as collaborators cite the institution more than non-collaborators; however, the influence of collaborators decreases as the institution's impact decreases (Figures 3-8 & 10-12). Hence, on average, higher impact institutions gain more benefit from the collaborators than the lower-impact ones. They presumably produce the most important works in the mainstream science, so their collaborators prefer to cite them back and probably they are highly requested for collaboration rather than lower-impact institutions.

Figures 3-8 and 10-12 also show that on average, as the institutions' impact increases, the effect of collaborators on the citations' impact of the institutions increases. For instance, a comparison of Figure 3 and 5 show that on average, lower-impact institutions receive 19% less citations from their collaborators than the high-impact ones. This difference might be influenced by the number of collaborations because presumably high-impact institutions are more productive, so they have more collaborators and this increases the chance of being cited by the collaborator networks. Lower-impact institutions that are more collaborative receive more citations from their collaborators than what gained by high-impact institutions from their collaborators (Figures 3, 5, 6, 8, 10 and 12). Thus, it seems that being more collaborative is a good strategy for lower impact institutions; however, the collaborators of lower impact institutions highly cite their joint papers with the institution (Figures 6), probably because they have less similar research profile with the lower-impact institutions. Therefore, the results vary considering which strategy, balancing or choosing between being highly collaborative or producing highly impact research, is important for the institution. Figures 3-8 and 10-12 also reveal that the difference between middle and high collaborative institutions in the number of received citations from the collaborators is slighter than the difference between each of these two groups of institutions and lower collaborative institutions.

Given that high impact institutions contribute to the more important research they are likely to attract both collaborators and citations. In contrast, lower-impact institutions may be focusing on research topics that attract less attention, which might explain why their collaborators might be more likely to cite jointly written papers than papers to which they did not collaborate. That may also explain the fact that, on average, lower-impact institutions receive more citations from peripheral-citers group and also core-citers and non-collaborators groups compared to the higher-impact institutions (Figures 6).

On the whole, although institutions receive a higher credit from their collaborators, this effect varies according to a) the degree of institution's contribution to collaboration, b) institutions' impact level, c) different groups of collaborators including core and peripheral groups, d) different groups of citers including core and peripheral groups, e) publication year of the institution's cited documents, and f) different groups of publications including collaborated papers with the citing institutions and non-

collaborated ones. The effect of collaboration decreases as the institutions level of impact increases. Hence, collaborating more does not directly imply obtaining higher impact.

References

- Ahn, J., Oh, D. H., & Lee, J. D. (2014). The scientific impact and partner selection in collaborative research at Korean universities. *Scientometrics*, 100(1), 173-188.
- Ajiferuke, I., Lu, K., & Wolfram, D. (2011). Who are the research disciples of an author? Examining publication recitation and oeuvre citation exhaustivity. *Journal of Informetrics*, 5(2), 292–302.
- Andrade, H. B., de Los Reyes Lopez, E., & Martín, T. B. (2009). Dimensions of scientific collaboration and its contribution to the academic research groups' scientific quality. *Research Evaluation*, 18(4), 301-311.
- Bornmann, L., & Daniel, H. D. (2010). The citation speed index: A useful bibliometric indicator to add to the h index. *Journal of Informetrics*, 4(3), 444-446.
- Chang, H. W., & Huang, M. H. (2013). Prominent institutions in international collaboration network in astronomy and astrophysics. *Scientometrics*, 97(2), 443-460.
- Defazio, D., Lockett, A., & Wright, M. (2009). Funding incentives, collaborative dynamics and scientific productivity: Evidence from the EU framework program. *Research Policy*, 38(2), 293-305.
- Didegah, F., & Thelwall, M. (2013). Which factors help authors produce the highest impact research? Collaboration, journal and document properties. *Journal of Informetrics*, 7(4), 861-873.
- Gazni, A., & Didegah, F. (2011). Investigating different types of research collaboration and citation impact: a case study of Harvard University's publications. *Scientometrics*, 87(2), 251-265.
- Gazni, A., & Thelwall, M. (2014). The long-term influence of collaboration on citation patterns. *Research Evaluation*, 23(3), 261–271. doi: 10.1093/reseval/rvu014
- Gazni, A., & Thelwall, M. (2015). The citation impact of collaboration between top institutions: A temporal analysis. *Research Evaluation*, rvv039.
- Gazni, A., Sugimoto, C. R., & Didegah, F. (2012). Mapping world scientific collaboration: authors, institutions, and countries. *Journal of the American Society for Information Science and Technology*, 63(2), 323-335.
- Glänzel, W., Thijs, B. (2004) The influence of author self-citations on bibliometric macro indicators. *Scientometrics* 59(3), 281-310
- Goldfinch, S., Dale, T., & DeRouen, K. (2003). Science from the periphery: Collaboration, networks and 'Periphery Effects' in the citation of New Zealand Crown Research Institutes articles, 1995-2000. *Scientometrics*, 57(3), 321-337.
- Jones, B. F., Wuchty, S., & Uzzi, B. (2008). Multi-university research teams: shifting impact, geography, and stratification in science. *science*, 322(5905), 1259-1262.
- Larivière, V., Gingras, Y., Sugimoto, C. R., & Tsou, A. (2015). Team size matters: Collaboration and scientific impact since 1900. Forthcoming in the *Journal of the Association for Information Science and Technology*.
- Laudel, G. (2002). What do we measure by co-authorships? *Research Evaluation*, 11(1), 3-15.
- Levitt, J. M., & Thelwall, M. (2009). Citation levels and collaboration within library and information science. *Journal of the American Society for Information Science and Technology*, 60(3), 434-442.

- Levitt, J. M., & Thelwall, M. (2010). Does the higher citation of collaborative research differ from region to region? A case study of Economics. *Scientometrics*, 85(1), 171-183.
- Martin, T., Ball, B., Karrer, B., & Newman, M. E. J. (2013). Coauthorship and citation in scientific publishing. *arXiv preprint arXiv:1304.0473*.
- Thagard, P. (1997). Collaborative knowledge. *Noûs*, 31(2), 242-261.
- Thijs, B., & Glänzel, W. (2010). A structural analysis of collaboration between European research institutes. *Research Evaluation*, 19(1), 55-65.
- Van Dalen, H. P., & Henkens, K. N. (2005). Signals in science-On the importance of signaling in gaining attention in science. *Scientometrics*, 64(2), 209-233.
- Wallace, M. L., Larivière, V., & Gingras, Y. (2012). A small world of citations? The influence of collaboration networks on citation practices. *PloS ONE*, 7(3), e33339.
- White, H.D., Wellman, B., & Nazer, N. (2004). Does citation reflect social structure?: Longitudinal evidence from the “Globenet” interdisciplinary research group. *Journal of the American Society for Information Science and Technology*, 55(2), 111–126.
- Wray, K. B. (2002). The epistemic significance of collaborative research. *Philosophy of Science*, 69(1), 150-168.
- Yan, E. and Ding, Y. (2012) Scholarly Network Similarities: How Bibliographic Coupling Networks, Citation Networks, Cocitation Networks, Topical Networks, Coauthorship Networks, and Coword Networks Relate to Each Other, *Journal of the American Society for Information Science and Technology*, 63/7: 1313–26.
- Zhao, Q., & Guan, J. (2011). International collaboration of three ‘giants’ with the G7 countries in emerging nanobiopharmaceuticals. *Scientometrics*, 87(1), 159-170.