Scale-Adjusted Metrics of Scientific Collaboration

Éric Archambault[1], Olivier H. Beauchesne[2], Grégoire Côté[3], and Guillaume Roberge[4]

[1] *eric.archambault@science-metrix.com*

[2] *olivier.hbeauchesne@science-metrix.com*

[3] *gregoire.cote@science-metrix.com*

[4] *guillaume.roberge@science-metrix.com*
Science-Metrix, 1335 avenue du Mont-Royal E, Montréal, Québec, H2J 1Y6 (Canada)

Abstract
Scientific collaboration is increasing on nearly all fronts. In most fields of inquiry, the proportions of multiple authors', multiple institutions', and multiple countries' papers have increased regularly since the birth of scientific journals. Two questions that are frequently asked are: how does collaboration compare from one place to the other, and how does the intensity of collaboration between partners compare in systems with multiple players? For obvious reasons, absolute numbers do not reveal much, but it has been known since the 1970s that the percentages of collaboration present an inverse relationship relative to the number of papers. This paper presents scale-independent methods to examine how frequently collaboration occurs as a function of size. In addition to these scale-adjusted statistics, which are based on the use of the Katz normalization method, this paper proposes a new method to compute a scale-adjusted preference index of collaboration between entities of various sizes. Examples are provided for the world, the European Research Area (ERA), and the US states, as well as for Canadian universities.

Introduction
Since World War II, a notable shift has occurred regarding policy-makers' emphasis on the role played by collaboration in science. During the Cold War, the key word was 'secrecy'. Secrecy and competition have not disappeared yet, but neither has the need of countries—even those who are foes—to cooperate in many respects, and cooperation in science has increasingly played an important role for all but the most isolated countries. In fact, science presents yet another case of activity that combines the seemingly contradictory terms 'cooperation' and 'competition' (Colman, 1982). At the micro level, it has been suggested that the frequently observed independence, self-sufficiency, and self-directedness of scientists are balanced by the cooperative and communal character of science, which fosters teamwork and interdependence, which in turn foster collaboration (Fox & Faver, 1984).
Scientific collaboration has been studied broadly (Katz & Martin, 1997) and has been the object of discussion and of varying levels of empirical inquiry for a relatively long time (Bok, 1955; Bush & Hattery, 1956; Storer, 1970). In recent years, one of the most important catalysts for the growth of international collaboration has no doubt been the advent of big science—in particular, the construction of large accelerators, which among other factors has played a role in the development of a European research area (Heisenberg, 1973). Importantly though, the very presence of collaboration is due to certain *a priori* characteristics of science that require interpersonal sharing and reaching a general level of agreement on concepts and ideas (Apel, 1988); in fact, one can argue that all meaning issues forth from collaboration (Gusdorf, 1952).
Collaboration has been a constitutive aspect of science from its very beginning, as it was one way to transmit and improve knowledge. One can think of the flow of knowledge from Socrates to Plato and then to Aristotle. The increased mobility of scientists and the development of the postal system in the Renaissance certainly helped to increase collaboration through the exchange of ideas—one can think of the correspondence of Kepler

with other astronomers. Yet, collaboration in the age of the postal system—like that in the internet age—still depends on closeness. Kepler eventually moved from the University of Graz in Austria to work at the observatory of Tycho Brahe near Prague, and still today researchers are taking positions at attractive institutions, spending time as visiting scholars, and attending meetings at symposia. The growth of scientific journals and that of co-authorship, which have come about concomitantly with the professionalization of science and the development of increasingly sophisticated scientific instruments, have all helped to greatly accelerate the pace of collaboration (Beaver & Rosen, 1978). The growth of scientific societies is also an important factor to consider (Kerwin, 1981), although this is likely to be more of a corollary than a cause of a growing volition for collaboration. In the last 30 years, we have learned that scientific collaboration—more particularly international collaboration—translates into greater scientific quality (Presser, 1980), visibility (Beaver & Rosen, 1979; Bordons, 1996), and impact (Frenken, Holzl, & de Vora, 2005; Glänzel & Schubert, 2001; Narin, Stevens, & Whitlow, 1991; Smart & Bayer, 1986), and that international collaboration presents particular benefits to less countries (Glänzel, Schubert, & Czerwon, 1999). The effect of collaboration on scientific productivity should be measured carefully; it has been suggested that fractional counting methods are used (Aksnes, Rørstad, Sivertsen, & Piro, 2010; Archambault & Larivière, 2010; Braun, Glänzel, & Schubert, 2001), and the evidence obtained so far suggests that gains vary considerably across disciplines (Abramo, D'Angelo, & Di Costa, 2009).

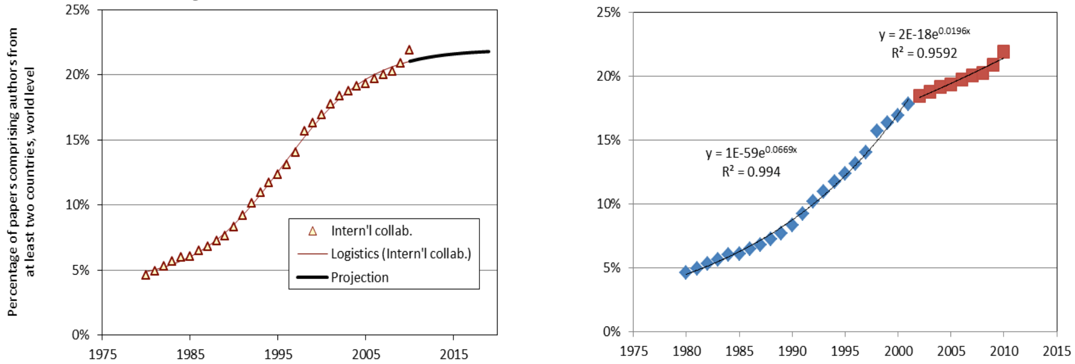
Collaboration has been studied for a long time using time-series (de Solla Price, 1963). Numerous studies have shown a regular increase in the propensity of individuals, institutions, countries, and other entities to collaborate. However, Figure 1 A suggests that the rate of international collaboration may have entered a steady state typically observed in S-shaped type of logistic growth. However, if one uses the model presented in Figure 1 B instead, perhaps we are only observing a momentary slowdown in the wake of the events of 9/11 and the increased barriers to cooperation that ensued. If the latter is the case, the system would be following what de Solla Price (1963) called a process of escalation.

More recently, scientists who borrowed methods from statistical physics have searched for universalities both in the topology of scientific networks and in the dynamics governing their evolution. Barabási et al. (2002) identified three important results stemming from this approach: most networks exhibit a so-called small world property, which means that the separation between nodes is small, the degree of clustering is higher than expected for random networks, and many networks follow a power-law distribution. Some of the findings in network analyses could have been expected considering what was shown in scientometrics long ago using stationary analyses. For instance, one can see in Frame and Carpenter (1979) three decades ago and Schubert and Braun (1990) two decades ago that the larger a country's research effort is, the smaller the proportion of international co-authorships associated with that country. More recently, Katz (2000) showed a power-law relationship between the number of publications by institutions and their number of collaborations.

The fact that collaborative activities are greatly influenced by scale is an intuitive proposition (Narin et al., 1991; Schubert & Braun, 1990). For instance, if there were only two countries in a system, with one country comprising a single researcher and the other a very large number of researchers, the smaller country would need to turn to the larger for every collaboration, whereas the proportion of external collaborations of the larger country would be very slight. It is therefore very important to take scale into account to determine whether entities collaborate more or less than expected.

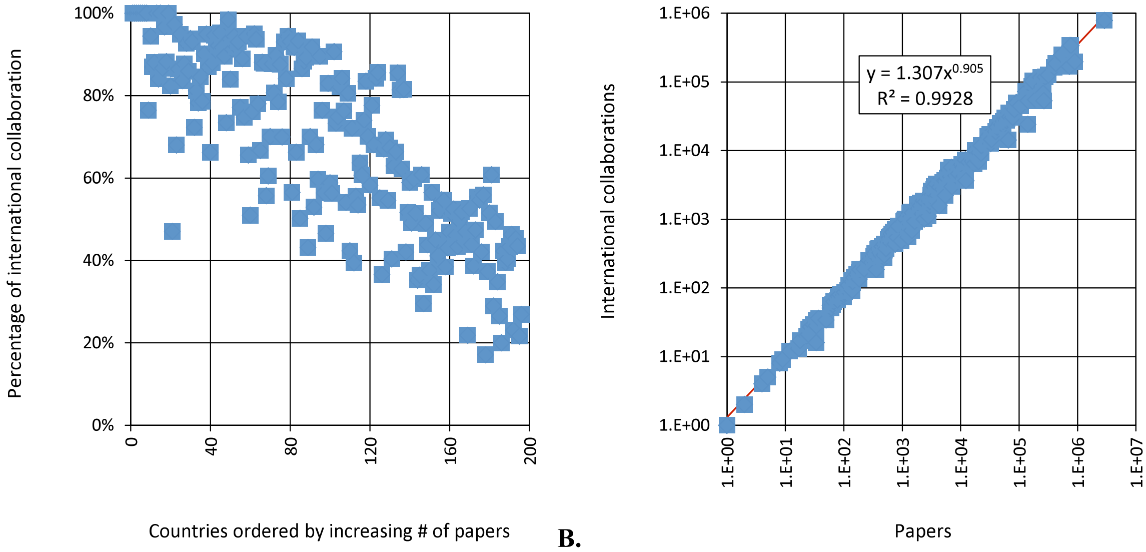
Another key question in the study of collaboration is how to determine whether an entity such as a country collaborates more or less than expected with another entity (Zitt, Bassecoulard, &

Okubo, 2000), and this question is relevant to entities of other scales as well such as states, provinces, and organisations.



A. B.

Figure 9. A. Logistic model, and B. Stage-model of international collaboration, 1980-2010



A. B.

Figure 10. A. Percentage of international collaboration, and B. Power-law relationship between number of papers and international collaborations, 2001-2010

There has been many attempts to date to determine the best way to calculate collaboration affinity between entities (Bookstein, Moed, & Yitzahki, 2006), including applications in collaboration networks (Elmacioglu & Dongwon, 2009; Eriksen & Hornquist, 2001) and attempts to describe (Newman, 2001) and explain the patterns of preferential attachments (Kim & Jo, 2010). In fact, just like the propensity of countries to collaborate following a power-law relationship between the number of publications and the number of international collaborations, the same kind of relationship has been found in real world networks. Real world networks differ from uniform random graphs with respect to node connectivity distribution. The power-law distribution in social networks differs from that of nodes in random Erdos-Renyi networks, which presents a Poisson law constant probability and degree distribution (Albert & Barabási, 2002; Barabási & Albert, 1999; Roth, 2005). Salton and McGill's (1983) measure has been widely used to examine collaboration affinity, but it presents a drawback because it is symmetrical. The Salton measure "is calculated for a pair of countries as the number of joint papers divided by the geometric mean (square root of the product) of the two countries' totals" (Glänzel & Schubert, 2001). Importantly however, just as love between humans is not always reciprocal, the attraction of countries, states, and institutions is not necessarily reciprocal and symmetrical either. Actually, although some can be singularly attracted to an entity, there are no doubt cases where the same entity actually feels some aversion (a repulsive force) towards its lover(s).

In order to alleviate this problem, Glänzel (2001) and Glänzel and Schubert (2001) proposed the use of a co-authorship affinity index, which Schubert and Glänzel (2006) later renamed the 'co-authorship preference index':

> *Co-authorship preference* of a country *i* towards country *j*, is measured by the coauthorship preference index, *QA*(*i,j*):
>
> $$QA(i,j) = (XA(i,j)/XA(i,*))/(XA(*,j)/(XA(*,*)–XA(i,*))),$$
>
> where *XA*(*i,j*) is the element in the *i*-th row and *j*-th column of the co-authorship matrix, *XA*(*i,**) and *XA*(**,j*) are the row sums of the *i*-th and *j*-th row, respectively (because of the symmetry of the co-authorship matrix they are equal to the corresponding column sums, as well), and *XA*(*,*) is the overall sum of the elements of the matrix. In words, the co-authorship preference index is the share of the joint publications of country *I* and country *j* in country *i*'s all international co-authorships divided by the share of country *j*'s all international co-authorships in world's total international co-authorships excluding those of country *i*.

The potential weakness of this approach is that it takes two measures that are highly influenced by scale (i.e., percentages of collaboration) and uses them in a ratio. It is important to state that simple ratios such as percentages should be used when there is a linear, random, relationship between the units. Thus, Glänzel and Schubert's (2006) method assumes that collaboration occurs in a random network, which contrasts with the evidence accumulated in the last 30 years.

This paper presents a generalisation of the scale-adjusted collaboration metrics to examine preferred patterns of collaboration between entities of various scales (such as countries, states, and universities) that reflect the non-linear relationships between published output and collaborations between entities. The method proposed here is simultaneously more comprehensive and more robust than previous indicators of collaboration affinity, such as Salton's measure, because it provides asymmetric measures and it addresses—directly, parsimoniously, and quite elegantly—the need to reflect the presence of the power-law relationship between size and collaboration affinity. Because of the fractal, self-similar character of systems describe by power-law, it appeared relevant to use the approach in systems of various size, which explains why examples are provided for the world, the European Research Area (ERA), the US states, and Canadian universities.

Methods

The method used here is that suggested by Katz (2000). It involves calculating an expected value based on a power-law regression of the relationship between two related variables. In our case, the variables are the number of papers and the number of collaborations. As shown time and again, a power-law relationship exists between the number of collaborations (*C*) and the number of papers (*P*):

$$C \sim bP^{a}$$

where *b* is a constant and *α* is the scaling coefficient obtained by regression. In this case, using the normalisation method proposed by Katz, the level scale-adjusted collaboration intensity of an entity *e* ($SACI_e$) is given by:

$$SACI_e = \frac{C_e}{\hat{C}_e} = \frac{C_e}{bP_e^{\alpha}}$$

where $\hat{C}_e$ is the estimated number of collaborations. In the case of the construction of a matrix of scale adjusted collaboration affinity in a self-organised system comprising *k* entities ($SaACA_{e(1\ldots k)}$), the matrix is calculated column by column by regressing the relationship between the number of papers (*P*) of every entity (e.g. papers entity 2, papers entity 3, … , papers entity *k*) with the number of collaborations of each of these entities with the entity examined for each column (e.g. collaborations of entity 1 with entity 2, collaborations of entity 1 with entity 3, … , collaborations of entity 1 with entity *k*):

$$
\begin{array}{ccccc}
 & \boldsymbol{e_1} & \boldsymbol{e_2} & \cdots & \boldsymbol{e_k} \\
\boldsymbol{e_1} & — & \dfrac{C_{e_{2\leftrightarrow 1}}}{bP_{e_1}^{\alpha_2}} & \cdots & \dfrac{C_{e_{k\leftrightarrow 1}}}{bP_{e_1}^{\alpha_k}} \\
\boldsymbol{e_2} & \dfrac{C_{e_{1\leftrightarrow 2}}}{bP_{e_2}^{\alpha_1}} & — & \cdots & \dfrac{C_{e_{k\leftrightarrow 2}}}{bP_{e_2}^{\alpha_k}} \\
\cdots & \cdots & \cdots & — & \cdots \\
\boldsymbol{e_k} & \dfrac{C_{e_{1\leftrightarrow k}}}{bP_{e_k}^{\alpha_1}} & \dfrac{C_{e_{2\leftrightarrow k}}}{bP_{e_k}^{\alpha_2}} & \cdots & —
\end{array}
$$

Results

Collaboration Intensity

When one uses percentage of collaboration, it is difficult to be able to assert whether countries have a high propensity to collaborate because the smaller the country, the more it collaborates. Here, we are presenting scale adjusted collaboration indices and we have taken the log of the value to produce a symmetrical indicator (so that an equal amount of affinity and non-affinity would have the same lengths on either the of the central bar of the graph which represents the scaling relationship in the system), where value below 0 means that countries collaborate less than expected and, conversely, where positive values mean that countries collaborate more than expected considering their level of scientific production. Data extracted from the Web of Science (computed by Science-Metrix) is presented only for the 19 countries composing the G20 but was calculated on the basis of the output of all countries. One can see in Figure 3 that Indonesia, Germany, France, the UK, Canada and Italy are actively engaged in international scientific collaboration. Conversely, Turkey and Asian countries (India, China, Japan, Korea), American countries (Brazil, US, Argentina, Mexico) as well as Russia and South Africa, are not collaborating as much as could be expected given their size.

Figure 4 presents data from Scopus (computed by Science-Metrix) for the intra-European Research Area collaboration. Not surprisingly, Luxembourg and Belgium collaborate more than expected by their size which can likely be explained by the presence of European institutions. Many of the late accessing countries which not so long ago where located on the other side of the Iron Curtain are still not collaborating as much as their size would lead to expect (Lithuania, Slovenia, Poland, Romania) but not all countries that do not have relatively intensive collaboration with the ERA countries are among this group. For example, Greece, Malta, the UK, Spain, Italy and Germany collaborate less than expected. Conversely, Bulgaria, Slovakia, and Hungary have all embraced the ERA.

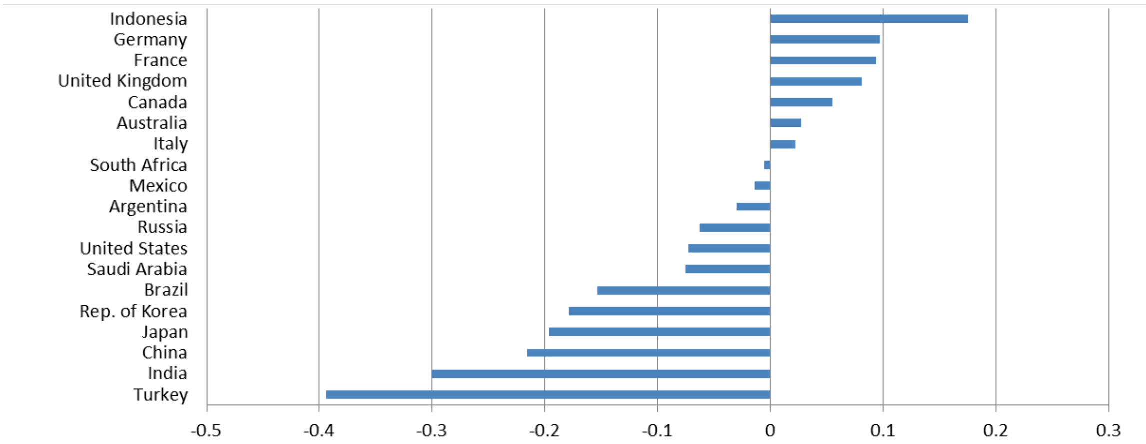


Figure 3. Collaboration intensity of the 19 G20 countries, 2001-2010 (Web of Science)

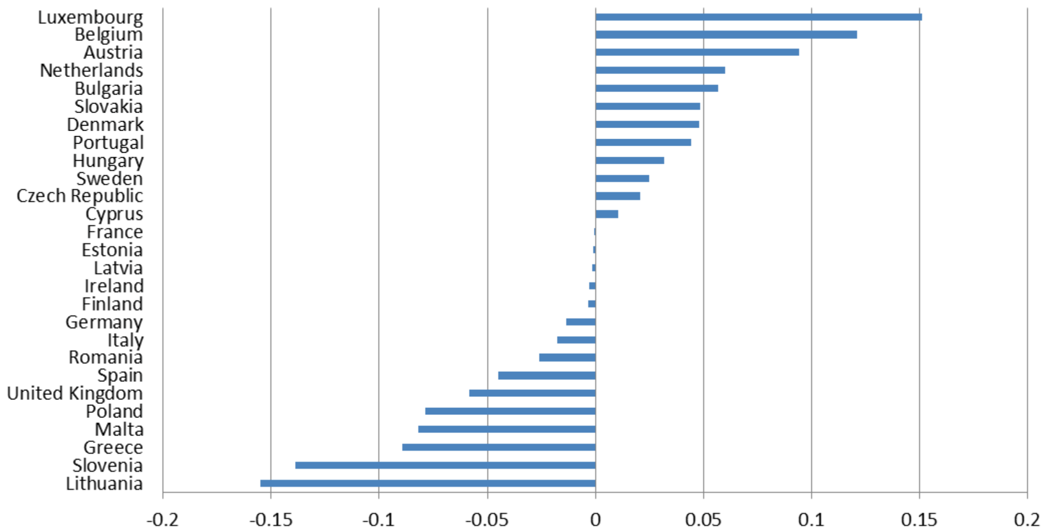


Figure 4. Intra-European Research Area collaboration intensity, 2000-2009 (Scopus)

Collaboration Affinity

Matrix obtained based on the method presented in the methods section were used to produce collaboration affinity maps in Gephi. Here, the size of the nodes represents the number of papers and the length of the edges is the result of the layout calculated with the force atlas algorithm. Only the edges showing the strongest affinities are shown in order to avoid producing graphs covered by edges (for instance, the total number of edges for the world countries map would be ~ 200^2).

Figure 5, which presents data from Scopus (computed by Science-Metrix), reveals a complex web of affinity shaped by historical variables such as colonial history (e.g. France being close to its ex-colonies in Africa, likewise for the UK as well as for Belgium and even for Portugal which is close to Angola and to Cape Verde). Geographic and linguistic proximity also play an important role as can be seen on the lower right side which comprises Latin American countries. Geographic, cultural and religious proximity certainly help explain the Muslim country partition on the left-hand side of the map. Also note the proximity of four Magreb countries, closer to France than other Asian countries are in a partition in the centre-left part of the map and there is a large partition comprising Baltic, Balkan, Central Europe and Mediterranean countries at the bottom and lower central part of the map. Interestingly, the European Union 27-countries are scattered in many parts of the map instead of presenting a cohesive partition. Speaking of which, the collaboration affinity of the members of the

European Research Area is shown in Figure 6 (data from Scopus computed by Science-Metrix). Three partitions can be found: Baltic countries, Central Europe countries with Germany, Liechtenstein, and Switzerland, and Western European countries which also comprise Cyprus, Greece, Malta, and Israel.

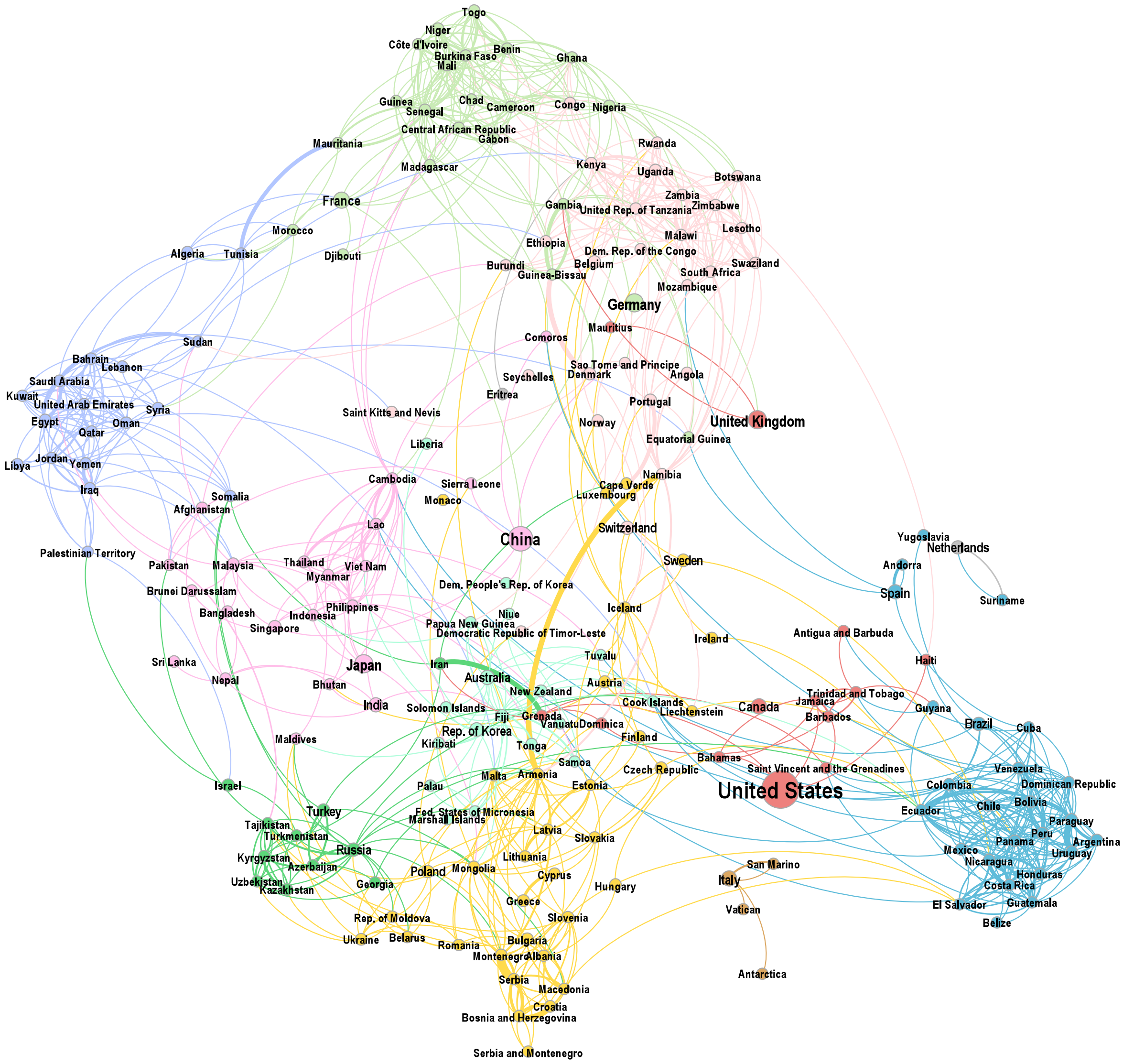


Figure 5. Affinity for international collaboration, 2003-2009 (Scopus)

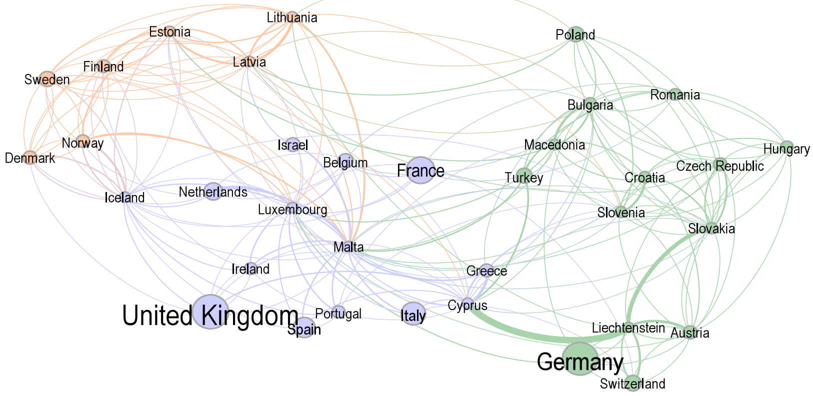


Figure 6. Affinity for intra-European Research Area collaboration, 2003-2009 (Scopus)

Figure 7 shows the US interstate collaboration affinity (data from the Web of Science computed by Science-Metrix). We obtained four partitions in Gephi, on the top-left group of states correspond to the western states (with the exception of the District of Columbia), the top right states in the graph correspond to the north-eastern states, the bottom left partition comprises the central states, while the bottom right partition comprises the south eastern states. This confirms that collaboration affinity is greatly shaped by geographic proximity.

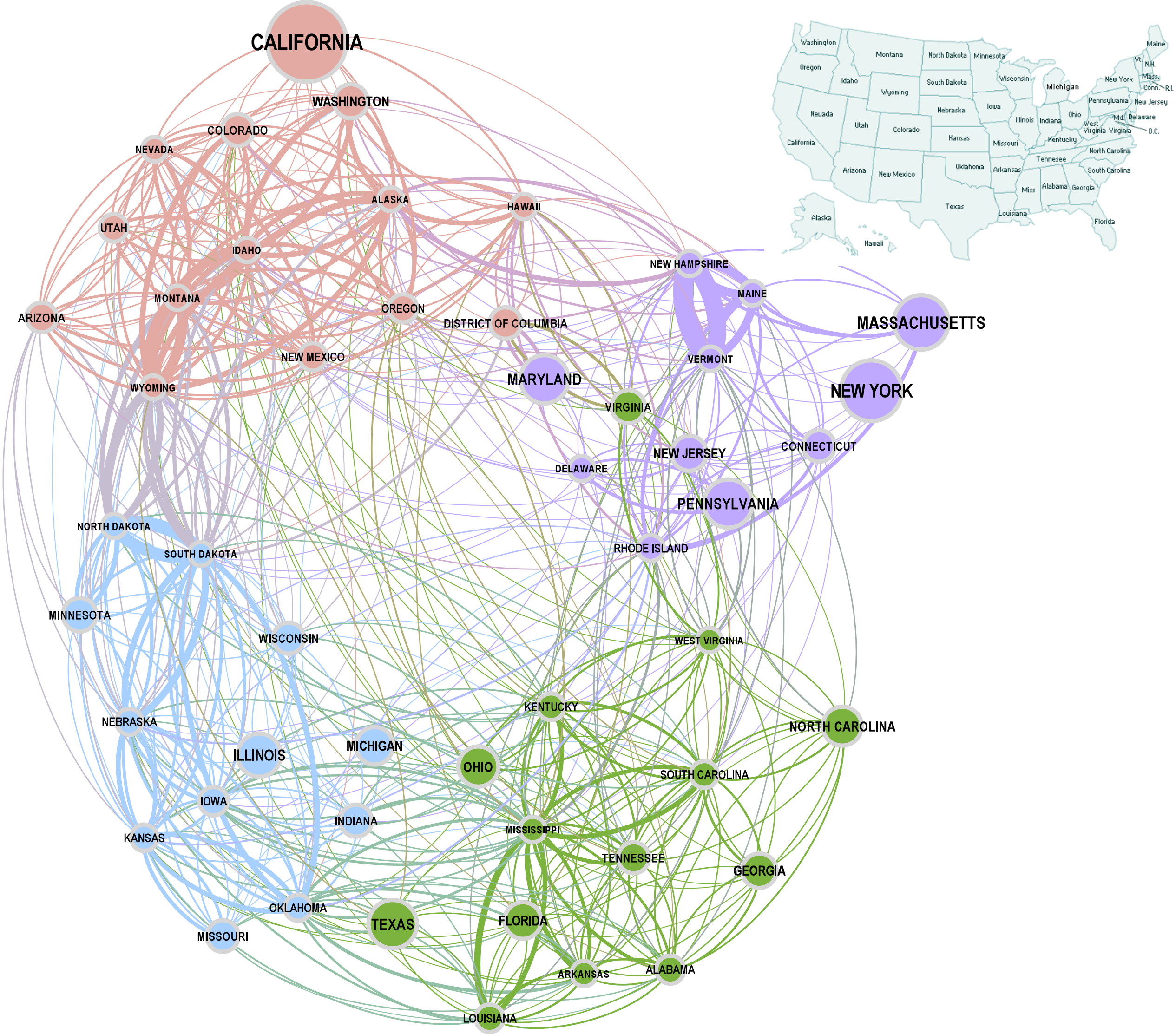


Figure 7. Affinity for Interstate collaboration, USA 2001-2010 (Web of Science)

Finally, Figure 8 presents the collaboration affinity of Canadian universities (using data from Scopus computed by Science-Metrix). Here, one can see four partitions where the main drivers for collaboration affinity clustering are a mix of geographical and language proximity. On the top left-hand side, one can mainly find Ontario universities, except for Carleton and Ottawa, which are located closer to Quebec. Though Carleton still belongs to the Ontario partition, Ottawa belongs to the Western Canada partition which can be found on the left hand side of the map. The Eastern Canada partition (and more agriculture-oriented universities such as Guelph) is one the top right part while Quebec and the mostly French speaking institutions (such as Moncton which is in New Brunswick) form a compact partition on the bottom part of the map.

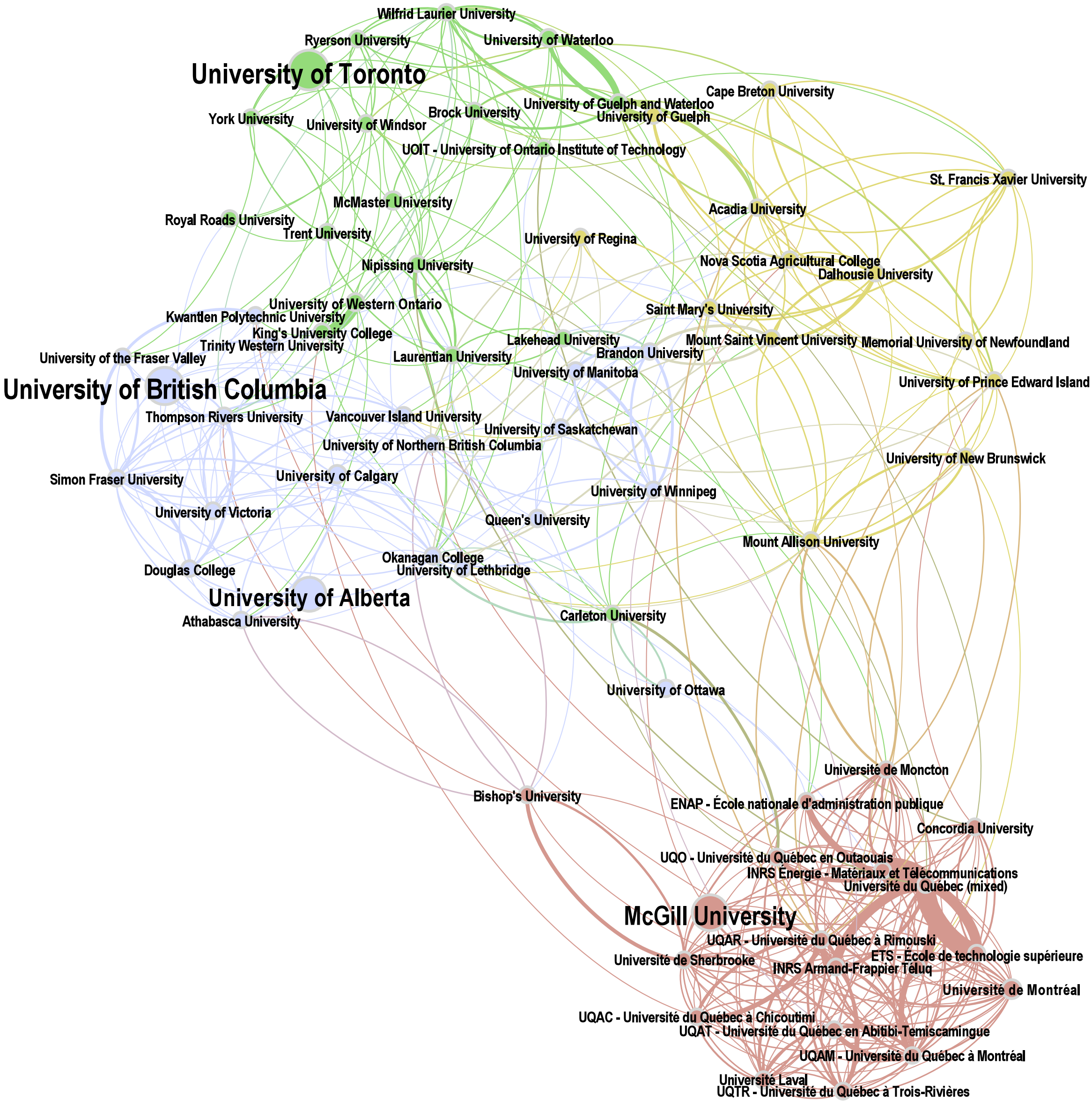


Figure 8. Affinity for inter-university collaboration, Canada, 1996-2009 (Scopus)

Conclusion

Though it has been known since the late 1970s that collaboration intensity follows a power-law relationship and that the larger an entity, the less it tends to collaborate intensely with outside partners, one than one researcher has continued to use percentages and ratios to address questions of research intensity and affinity. The discovery by Katz (2000) that one can derive performance indicators by dividing observed values by expected values calculated using a hyperbolic regression on the data mean that we now have a solution to address the non-linear properties of collaboration. The graphs shown here certainly suggest that this method is intuitively robust. The maps present patterns that can easily be explained in terms of geographical, cultural, linguistic and historical affinity between countries. This study shows that geography and history are in fact still very much present in today's scientific landscape.

References

Abramo, G., D'Angelo, C. A., & Di Costa, F. (2009). Research collaboration and productivity: is there correlation? *Higher Education, 57*(2), 155-171.

Aksnes, D., Rørstad, K., Sivertsen, G., & Piro, F. (2010). Productivity differences across fields: a macro analysis. Paper presented at STI Conference 2010*,* Leiden, Netherlands.

Albert, R., & Barabási, A. L. (2002). Statistical mechanics of complex networks. *Reviews of Modern Physics,* 74, 47–97.

Apel, K. O, (1988). Understanding and explanation: A transcendental-pragmatic perspective. Cambridge: MIT Press.

Bok, B. J. (1955). Science in international cooperation. *Science, 121*(3155), 843-847.

Archambault, E., & Larivière, V. (2010). Individual researchers' research productivity: A comparative analysis of counting methods. Paper presented at STI Conference 2010*,* Leiden, Netherlands.

Barabási, A. L. & Albert, R. (1999). Emergence of scaling in random networks. *Science, 286,* 509–512.

Barabási, A. L., Jeong, H., Neda, Z., Ravasz, E., Schubert, A., & Vicsek, T. (2002). Evolution of the social network of scientific collaborations. *Physica A., 311,* 590-614.

Beaver, D. D., & Rosen, R. (1978). Studies in scientific collaboration: PART I. The Professional Origins of Scientific Co-Authorship. *Scientometrics, 1*(1), 65-84.

Bookstein, A., Moed, H., & Yitzahki, M. (2006.) Measures of international collaboration in scientific literature: Part I & II. *Information Processing and Management, 42,* 1408-1427.

Bordons, M. I., Gomez, I., Fernandez, M. T., Zulueta, M. A., & Mendez, A. (1996). Local, domestic and international scientific collaboration in biomedical research. *Scientometrics, 37*(2), 279-295.

Braun, T., Glänzel, W., & Schubert, A. (2001). Publication and cooperation patterns of the authors of neuroscience journals. *Scientometrics, 51*(3), 499-510.

Bush, G. P., & Hattery, L. H. (1956). Teamwork and creativity in research. *Science Quarterly, 1,* 361-362.

Colman, A. M. (1982). *Cooperation and competition in humans and animals*. Wokingham: Van Nostrand Reinhold.

Crane, D. (1972). Transnational networks in basic science. In R. O. Keohane & J. S. Nye (Eds), *Transnational Relations and World Politics* (pp. 235-251). Cambridge: Harvard University Press.

de Solla Price, D. J. (1963). Little science, big science. New York: Columbia University Press.

Elmacioglu, E., & Dongwon, L. (2009). Modeling idiosyncratic properties of collaboration networks revisited. *Scientometrics, 80*(1), 197-218.

Eriksen, K. A., & Hornquist, M. H. (2001). Scale-free growing networks imply linear preferential attachment. *Physical Review E, 65*(1), 017102.1-017102.2.

Fox, M. F., & Faver, C. A. (1984). Independence and cooperation in research: The motivations and costs of collaboration. *The Journal of Higher Education, 55*(3), 347-359.

Frame, J. D., & Carpenter, M. P. (1979). International research collaboration. *Social Studies of Science, 9*(4), 481-497.

Frenken, K., Holzl, W., & de Vora, F. (2005). The citation impact of research collaborations: The case of European biotechnology and applied microbiology (1988–2002). *Journal of Engineering and Technology Management, 22*, 9-30.

Glänzel, W. (2001). National characteristics in international scientific co-authorship relations. *Scientometrics, 51*(1), 69-115.

Glänzel, W., & Schubert, A. (2001). Double effort=double impact? A critical view at international co-authorship in chemistry. *Scientometrics, 50*(2), 199-214.

Glänzel, W., Schubert, A. & Czerwon, H.J. (1999). A bibliometric analysis of international scientific cooperation of the European Union (1985-1995). *Scientometrics, 45*(2), 185-202.

Gusdorf, G. (1952). *La Parole*. Paris: Presses Universitaires de France.

Heisenberg, W. (1973). Construction of large accelerators: Scientific and political aspects. *Science. 179*(4074), 643-647.

Katz, J. S. (2000). Institutional recognition: Scale-independent indicators and research. *Science and Public Policy, 27*(1), 23-36.

Katz, J. S. & Martin, B. R. (1997). What is research collaboration? *Research Policy, 26,* 1-18.

Kerwin, L. (1981). International science—an overview. *Science, 213*(4512), 1069-1072.

Kim, J. Y., & Jo, H. H. (2010). A signalling explanation for preferential attachment in the evolution of social networks. *Journal of Evolutionary Economics, 20,* 375-393.
Narin, F., Stevens, K. & Whitlow, E.S. 1991. Scientific co-operation in Europe and the citation of multinationally authored papers. *Scientometrics*. 21(3): 313-323.
Newman, M. E. J. (2001). Clustering and preferential attachment in growing networks. *Physical Review E*., 64, 025102.
Presser, S. (1980). Collaboration and the quality of research. *Social Studies of Science, 10*(1), 95-101.
Roth, C. (2005). Generalized preferential attachment: Towards realistic socio-semantic network models. *ISWC 4th International Semantic Web Conference, Workshop on Semantic Network Analysis, 171*, 29-42.
Salton, G., & McGill, M. J. (1983). *Introduction to modern information retrieval*. New York: McGraw Hill.
Schubert, A. & Braun, T. (1990). International collaboration in the sciences, 1981-1985. *Scientometrics, 19*(1-2): 3-10.
Schubert, A. & Glänzel, W. (2006). Cross-national preference in co-authorship, references and citations. *Scientometrics, 69*(2), 409-428.
Smart, J. C., & Bayer, A. E. (1986). Author collaboration and impact: A note on citation rates of single and multiple authored articles. *Scientometrics, 10*(5-6), 297-305.
Storer, N. J. (1970). The internationality of science and the nationality of scientists. *International Social Science Journal, 22,* 89-104.
Zitt, M., Bassecoulard, E., & Okubo, Y. (2000). Shadows of the past in international cooperation: Collaboration profiles of the top five producers of science. *Scientometrics, 47*(3), 627-657.