# **Supplementary Material**

**S.1 Data**

Drawing on the Web of Science (WOS) database, we searched for all publications concerning robotic surgery according to the following search query:

* Topic*: “robotic OR robotic OR robot assisted OR robotic assisted OR robotically assisted OR robot-assisted OR robotic-assisted OR robotically-assisted”*;
* Research Area*: “Surgery”*;
* Document Types*: “Article”*.

The search was performed on the 17th of January 2017, and produced 3,889 publications (peer-reviewed articles) published between July 1988 and January 2017. Figure S1 shows the distribution of the retrieved publications across the years of retrieved.

Of the 3,889 articles collected, 23 did not specify the affiliations of the co-authors. Thus, we focused on the remaining 3,866 articles with specified affiliations of co-authors. Based on these 3,866 articles, a process of manual disambiguation of name variants of organizations was performed. Out of 2,590 name variants, we filtered out 1,700 unique names. To this end, each name variant was searched for on the WOS “Organizations - Enhanced List” tool with a view to identifying the preferred name associated with the corresponding organization. If such name was found, then it was used as a unique name. In particular, 816 preferred names were found corresponding to 1,485 name variants. Not all organizations in the WOS database were included in this list by the WOS team. For 1,105 name variants we could not find a corresponding preferred name using the above mentioned tool. A manual search on Google search was then performed that took into account the associated geographical location (country and city) of the organization and, when necessary, also the affiliated authors’ names. This resulted in the identification of additional 884 unique names of organizations. Thus, in total we obtained 1,700 unique organizations.

Moreover, these 1,700 organizations were classified into 642 academic institutions, 11 academic systems, 863 health providers, 99 enterprises, 53 research institutes, 22 governmental entities, 8 research foundations, and 2 research working groups. These institutional categories were produced partly based on the WOS *Organization Type* classification produced through the InCites© (Clarivate Analytics, London, UK) data intelligence tool, and partly manually (mainly for enterprises and health providers that were not included in InCites).

The 1,700 organizations were nested within 62 countries, and within 6 geographical regions. These regions were based on the standard international allocation of countries into continents, with the exception of Russia and Middle Eastern countries that were classified as transcontinental countries. In particular, Russia was classified as belonging to Europe given the geographical proximity of its capital to the European borders.



**Figure S1.** Time distribution of publications

For each of the top 20 countries ranked by number of publications, Table S1 shows the number of organizations contributing to research.

|  |  |  |
| --- | --- | --- |
| **Country** | **Number of publications** | **Number of organizations** |
| USA | 1846 | 520 |
| Germany | 359 | 137 |
| South Korea | 352 | 58 |
| Italy | 280 | 127 |
| Japan | 224 | 112 |
| France | 204 | 103 |
| China | 179 | 71 |
| United Kingdom | 167 | 68 |
| Canada | 159 | 33 |
| Netherlands | 88 | 32 |
| Austria | 86 | 20 |
| Turkey | 79 | 49 |
| Belgium | 61 | 17 |
| Spain | 46 | 41 |
| Switzerland | 40 | 14 |
| Australia | 35 | 31 |
| Brazil | 34 | 26 |
| Taiwan | 34 | 22 |
| Romania | 30 | 17 |
| India | 27 | 18 |

**Table S1.** Top 20 countries ranked by number of published articles.

**S.2 The network**

We built up the collaboration network (see Figure 2 in the main manuscript) among the 1,700 organizations that carried out research on robotic surgery. The nodes of the network are the organizations, and links refer to collaborations between organizations. Specifically, a link exists between any two organizations when one or more scholars affiliated with one organization co-authored at least one article with one or more scholars affiliated with the other organization. Out of the 1,700 organizations, 1,543 were connected through at least one collaboration. The largest connected component includes 1,334 organizations, and in total the network comprises 6,000 collaboration links.

To construct the network, we first define as the undirected and binary bipartite authorship matrix defining connections between each organization and each publication such that element if and only if organization contributed to article , and otherwise (see Figure S2.a).1 We then define as the undirected and weighted projected one-mode collaboration matrix, where each element if and only if organization collaborated with the organization . That is, when organization *i* and organization *j* co-authored at least one publication (see Figure S2.b).

The weight of each collaborative tie between any two organizations is measured according to the method proposed by Newman (2001).2 In particular, the total weight of the collaborative tie between two organizations *i* and *j* is assumed to be equal to the sum of the strengths contributed by each of the articles co-authored by the two organizations, where strength is defined as:

and is the number of organizations appearing in the byline of article *a*. In other words, the total weight of the collaborative tie between organizations *i* and *j* is (i) proportional to the number of articles these organizations co-authored; and (ii) inversely proportional to the number of organization involved in each article. (see Figure S2.d). Note that all articles published by one single organization were excluded from the above measure as they did not contribute to the co-authorship network.

****

**Figure S2.** Authorship network, collaboration network, full and fractional counting method for computing weights of collaborative ties. a) The bipartite network connecting organizations to publications; b) The weighted undirected one-mode collaboration network among organizations; c) Weights of links and node strength (i.e., the sum of the weights of the links incident upon the node) computed according the full counting method; d) Weights of links and node strength computed according to the fractional counting method (Newman, 2001).2 Notice that in the case of fractional counting, node strength is equal to the number of articles the corresponding organization published.

**S.3 Measures**

In this Section, we provide details on various dependent and independent measures used in the analysis. These measures include: innovativeness; research impact; the clustering coefficient; efficiency; closeness centrality; geographical dispersion; and strength of industrial collaboration.

**S.3.1 The innovation index**

We classified each article in terms of its level of evidence following the procedure proposed by Garas et al. (2017).3 Each articles was associated with one out of 8 categories corresponding to distinct surgical innovation stages as shown in Table S2.

|  |  |
| --- | --- |
| **Surgical Innovation Stage** | **Description** |
| **1** | Randomized controlled trial |
| **2** | Non-randomized controlled trial |
| **3** | Observational study with controls |
| **4** | Observational study without controls |
| **5** | First-in-human study |
| **6** | Cadaveric study |
| **7** | Animal study |
| **8** | Description of idea / laboratory evaluation |

**Table S2**. Description of surgical innovation stages with which articles can be associated according to their level of evidence.

We could assign level of evidence to 3,389 articles. Figure S.3 shows the temporal distributions of these articles in each evidence-based category.



**Figure S3.** Number of articles per level of evidence over time.

We then computed the *innovation index* following the definition given in Garas *et al*. (2017).3 For each organization *org*, the innovation index can be defined as

where is an integer value labelling the ordinal category associated with the implementation stage according to level of evidence, is the count of publications of organization *org* associated with category , and is the total number of publications associated with category across all organizations in the data set (i.e., in the set *G*). Figure S4 shows the ranking of countries according to the average of innovation indices of all the organizations within their borders. The figure also highlights the most innovative organizations within and between countries.



**Figure S4**. Ranking of countries and organizations according to normalized innovation index. To obtain the normalized innovation index, we divided each organization’s innovation index by the maximum value observed in the data set. Countries have been ranked according to the sum of organizations’ normalized innovation index. Grey shaded areas correspond to distances between lower and upper quartiles (interquartile range). Darker shaded areas denote distances between lower quartile and median. Whiskers correspond to values within 1.5 times the interquartile range. To obtain the normalized innovation index, we divided each organization’s innovation index by the maximum value observed in the data set.

**S.3.2 Research impact**

For each organization, we measured its academic performance using the sum of the normalized citations received by all the article published by the organization across the years:

where refers to the normalized citations received by article published by (authors affiliated with) organization *org* in year *y*. To obtain normalized citations, the citation count for each publication in a given year was divided by the average number of citations obtained by all articles published in the same year:

where the average was computed across all articles published by all organizations in year *y*.

**S.3.3 The local clustering coefficient**

Network analysis provides a set of measures for quantifying a number of structural properties at different levels of a system.1 For example, measures can be calculated at a node’s level to detect its topological position (e.g., broker between otherwise disconnected partners) in the ego-centered network (i.e., the network including connections between the node and its partners as well as connections between these partners), or in the whole network (e.g., how close the node is to all other others in the network). Network analysis can therefore provide valuable tools for uncovering which nodes are central to the innovation process (‘knowledge catalysts’), which have a more peripheral role, and which act as connectors between otherwise unconnected groups (‘knowledge brokers’). By constructing and examining each node’s ego-centered network, we aim to detect the extent to which a node’s neighbors are also connected with each other, and in so doing we engage with current debates on social capital and the relative benefits of closed versus open social structures.1

Using the weighted network as defined above, we computed the local clustering coefficient of each organization, which measures the extent to which the organization’s ego-centered network is a closed collaborative structure.1-5 Formally, the local clustering coefficient of organization in a binary network is defined as the ratio between the number of actual triangles containing the organization and its neighbors, and the maximum possible number of such triangles, :

 .

In our study, we computed the clustering coefficient only for organizations with (i.e., 1,203 organizations). Moreover, we computed the generalized clustering coefficient to account for the (fractional) weight of links according to the definition proposed in Opsahl and Panzarasa (2009).4 In particular, we used the arithmetic mean as a method to obtain the weights of triangles and triplets, respectively in the numerator and denominator of the generalized clustering coefficient.

**S.3.5 Efficiency**

To measure an organization’s opportunities of brokerage, we also computed the organization efficiency, namely the degree to which the organization’s ego-centered network is non-redundant.6 Efficiency thus captures the extent to which an organization’s ego-centered network is an open structure. To measure efficiency, we first computed effective size, a measure originally introduced by Burt (1992) to capture the extent to which each of the first neighbors of a node is redundant with respect to the other neighbors. Following Latora *et al*. (2013),5 an organization *org*’s effective size, can be defined as

where is the degree of organization *org* (i.e., the number of organizations with which *org* collaborated), and is the local binary local clustering coefficient (i.e., the coefficient presented in S.3.3 computed using the unweighted network).

The efficiency of organization *org* (i.e., the normalized effective size) can now be computed as the ratio between the organization’s effective size and degree:

Notice that, as with the local clustering coefficient (Section S.3.3), efficiency was computed only for organizations with (i.e., 1,203 organizations).

**S.3.4 Closeness centrality**

Drawing on the weighed network defined above, we computed the closeness centrality of each organization. The closeness centrality of an organization is a function of the inverse of the shortest distances separating the organization from each of all other organizations in the network (i.e., the smallest number of links separating the organization from each of the others).1 Formally, the closeness centrality *lorg* of a given organization *org* is defined as:

where is the total number of organizations in the data set, and is the length of the shortest path between organization and organization . In other words, we measured the closeness centrality of organization *org* as the mean of the inverse distances from *org* to all other nodes in the network.

Once again, we used the generalized version of closeness centrality for undirected networks to account for the (fractional) weights of links according to the definition proposed in Opsahl *at al.* (2010).7 In particular, we set the alpha tuning parameter in the generalized measure equal to one (see Equation 8 in Opsahl et al. (2010)). In this case, the Dijkstra shortest paths were used, and the identification procedure of these paths relied simply on tie weights and disregarded the number of intermediary nodes along the paths.

**S.3.5 Geographical entropy**

In addition to the network-based measures described above, for each organization we measured: (i) the diversity of the geographical locations of collaborators; and (ii) the strength of collaboration with industrial partners.

To measure the geographical diversity of collaborators, we computed geographical entropy. To this end, let us define as the set of countries in which the collaborating organizations of a given organization *org* are located, i.e., . Formally, the geographical entropy of organization *org* can be defined as follows:

where is the fraction of organization *org*’s collaborative effort towards organizations located in country (i.e., the ratio between the sum of the weights of links between *org* and all organizations in country and the sum of the weights of links between *org* and all organizations in ). Notice that each weight is measured once again according to the method proposed by Newman (2001) (see Section S.2).

Given the set of countries in the data set and the norm of such set, an organization’s geographical entropy can range from zero (when all international collaborators are located in the same country) to *log* (when the international collaborators are uniformly distributed across countries, i.e., all countries except the organization’s own). Similarly, given the set of countries with which the organization *org* collaborates, entropy is maximized when the organization equally distributes its collaborative efforts across all countries in . That is, for each country , is maximized when each fraction of collaborative effort is equal to , where is the norm of the set .

Thus, geographical entropy enables us to assess organizations beyond the mere count of international collaborators, and to distinguish between organizations that, while having the same number of international collaborators, differ in terms of their geographical dispersion. This is important as some organizations may have a large number of international collaborators, yet all confined within the same geographical boundaries, whereas others may have few collaborators that nonetheless are widely geographically dispersed.



**Figure S5.** Representation of geographical dispersion. a) An organization org can collaborate with domestic or international organizations, and , respectively. Each of the international organizations can, in turn, be associated to the foreign country in which it is located. The size of each cell is proportional to the effort devoted by organization org towards country (i.e., proportional to the sum of weights of the links between org and all organizations in country ). b) The distribution of organization org’s efforts across foreign countries. The value of each is equal to the ratio between and the sum of org’s efforts towards all countries in , i.e., . Based on the distribution of , it is possible to compute the geographical entropy .

**S.3.6 Industrial collaboration**

To assess the association between research impact and innovation on the one hand, and collaborative links with industry on the other, for each organization we measured the strength of industrial collaborations. Indeed collaboration with industry represents a potentially relevant determinant of research performance that may not be captured by more traditional network metrics.8

The strength of an organization’s industrial collaborations was computed as the sum of the organization’s collaborative efforts towards industrial partners. Formally, the strength of industrial collaborations of a given organization *org* is defined as:

where denotes the weight of the collaborative link between organization *org* and organization *j*, and *I* represents the set of all organizations classified as “Industry”. Notice that each weight is measured once again according to the method proposed by Newman (2001) (see Section S.2).

**S.3.7 Other control variables**

We also controlled for a measure of volume, i.e., the ‘scholarly output’ in WOS, here referred to as ‘number of articles in WOS’. This was measured through the number of all articles published by each organization that the InCites**©** intelligence tool could retrieve in the WOS database. Note that this control variable refers to each organization’s overall scholarly output in WOS beyond the specific field of robotic surgery. This, of course, should not be confused with the number of publications in our data set that are linked to any given organization. Moreover, WOS does not track all the organizations we were able to detect in our set of publications (within robotic surgery). Out of the 1,700 organizations in our data set, only 814 could be associated with a total scholarly output in WOS.

**S.4 Statistical analysis**

Maximum-likelihood estimates of two hierarchical three-level random-intercept models were computed in which organizations were nested within countries, in turn nested within geographical regions. A random-intercept structure was combined with heteroskedastic level-1 residuals by letting the variances of these residuals be a function of the organization’s institutional type. For organization *org*, country , and geographical region *r*, we estimated the following three-level linear random-intercept models:

and

where are level-1 covariates (as defined in Section S.3) for organization *org* in country and region *r*, is a level-1 covariate measuring the number of all articles published by organization *org* in country and region *r* and recorded in the WOS database, is the dummy variable for the *k*-th institutional type of organization *org* in country and region *r*, to are fixed parameters, is the country-level (i.e., level-2) random intercept, is the region-level (i.e., level-3) random intercept, and is the organization-level (i.e., level-1) error term.8

We let all observed covariates for all organizations in region *r* be denoted by the matrix **.** It was assumed that the region-level random intercept has zero expectation and variance given the covariates . Similarly, we assumed that the country-level random intercept has zero expectation and variance given the random intercept and the covariates .That is, we assumed that: (i) there is zero correlation between covariates and the random intercept at the region level (i.e., level-3 exogeneity); there is zero correlation between covariates and the random intercept at the country level (i.e., level-2 exogeneity); there is zero correlation between random intercepts and across regions and countries; the variance of the random intercept at the region level is homoskedastic given the covariates; and (iv) the variance of the random intercept at the country level is homoskedastic given the covariates and the region-level random intercept.

Moreover, it was assumed that the level-1 error term has zero expectation, given the covariates and the random intercepts and . That is, we assumed that there is zero correlation between covariates and level-1 residuals (i.e., level-1 exogeneity), and zero correlation between level-1 residuals and both random intercepts and . However, unlike the two random intercepts, the level-1 residuals were not assumed to be homoskedastic. To this end, we combined a three-level random-intercept structure with heteroskedastic level-1 residuals by letting the variances of these residuals be a function of the organization’s institutional type , given the covariates and level-3 and level-2 random intercepts. That is, the conditional variance of the responses given the covariates and random intercepts, or the conditional variance of the total residual, depends on the value of the covariate , and the total residual is therefore heteroskedastic. Finally, the random effects at each level were assumed to be mutually independent and independent of the level-1 residual error term.

**S.5 Results**

For the sake of completeness, Tables S3-S8 report estimates for all parameters, including the remaining fixed effects (i.e., intercept and institutional types) as well as the estimated random-effect parameters (i.e., the standard deviations of the level-3 and level-2 random intercepts and of the level-1error term, ), , respectively). To check for robustness, we also estimated a number of additional three-level random-intercept models. First, we replaced the local clustering coefficient with efficiency, a measure that directly captures the extent to which an organization collaborates with non-redundant partners that are otherwise disconnected with each other. Second, we introduced a level-1 interaction term between (standardized) geographical entropy and the local clustering coefficient.

**S.5.1 Estimates from mixed-effects regression models of research impact**

Table S3 reports the full range of estimates from the hierarchical three-level random-intercept model of research impact including the local clustering coefficient as a covariate (as in the main manuscript).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 0.0015294 | 0.0021908 | 0.7 | 0.485 | -0.0027645 | 0.0058232 |
| **Clustering coefficient (*cc*)** | -0.40268 | 0.616509 | -0.65 | 0.514 | -1.61102 | 0.805651 |
| **Closeness centrality (*l*)** | 6.078229 | 0.725326 | 8.38 | 0 | 4.656616 | 7.499843 |
| **Geographical entropy ()** | 10.55992 | 0.756405 | 13.96 | 0 | 9.077395 | 12.04245 |
| **Industrial collaboration (*ic*)** | 5.210371 | 1.283682 | 4.06 | 0 | 2.6944 | 7.726341 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | -0.63742 | 2.850473 | -0.22 | 0.823 | -6.22424 | 4.949406 |
| **Industry** | -6.10235 | 1.589754 | -3.84 | 0 | -9.21821 | -2.98649 |
| **Government** | -6.98335 | 1.991627 | -3.51 | 0 | -10.8869 | -3.07983 |
| **Health** | -1.94396 | 1.272346 | -1.53 | 0.127 | -4.43771 | 0.549792 |
| **Research institute** | 2.056959 | 2.771084 | 0.74 | 0.458 | -3.37427 | 7.488183 |
|  |  |  |  |  |  |  |
| **Constant** | 1.641568 | 1.711969 | 0.96 | 0.338 | -1.71383 | 4.996965 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 1.754686 | 0.986096 | 0.583233 | 5.27906 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 1.645659 | 1.159483 | 0.413626 | 6.54744 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 20.37447 | 0.706987 | 19.03487 | 21.80834 |
| **Academic system** | 5.749711 | 1.93462 | 2.973309 | 11.11865 |
| **Industry** | 3.104733 | 0.894139 | 1.765569 | 5.459635 |
| **Government** | 5.184769 | 1.134586 | 3.376451 | 7.961564 |
| **Health** | 8.482241 | 0.567122 | 7.440451 | 9.669898 |
| **Research Institute** | 13.80572 | 1.863823 | 10.59604 | 17.98764 |
| Number of observations = 639Log pseudolikelihood = 2678.401Wald (10) = 458.82Log likelihood = -2678.401  |

**Table S3**. Maximum-likelihood estimates from three-level random-intercept linear models of normalized citations with the local clustering coefficient as covariate.

An increase by one standard deviation in (normalized) closeness centrality was associated with an expected increase of 6.0782units in normalized. After controlling for the other covariates, a one-unit increase in geographical entropy was associated with an expected increase in normalized citations by 10.5599 units. Estimates suggested that there was no significant difference in (normalized) citation count between two organizations that differed in number of publications, after controlling for the other covariates.

Table S4 reports the full range of estimates from the hierarchical three-level random-intercept model of research impact in which the local clustering coefficient has been replaced by efficiency. Results are in qualitative agreement with the ones reported in the main manuscript. A surge in opportunities of knowledge brokerage are associated with an expected increase in research impact, after controlling for the other covariates; yet the association is not statistically significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 0.001534 | 0.002197 | 0.7 | 0.485 | -0.00277 | 0.005841 |
| **Efficiency (*eff*)** | 0.509785 | 0.570665 | 0.89 | 0.372 | -0.6087 | 1.628268 |
| **Closeness centrality (*l*)** | 5.972611 | 0.727801 | 8.21 | 0 | 4.546148 | 7.399074 |
| **Geographical entropy ()** | 10.84422 | 0.774389 | 14 | 0 | 9.326449 | 12.362 |
| **Industrial collaboration (*ic*)** | 5.284805 | 1.290504 | 4.1 | 0 | 2.755464 | 7.814145 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | -0.66534 | 2.847907 | -0.23 | 0.815 | -6.24714 | 4.916454 |
| **Industry** | -5.93807 | 1.622925 | -3.66 | 0 | -9.11894 | -2.75719 |
| **Government** | -6.99423 | 1.933888 | -3.62 | 0 | -10.7846 | -3.20387 |
| **Health** | -1.91017 | 1.275366 | -1.5 | 0.134 | -4.40984 | 0.589505 |
| **Research institute** | 2.17482 | 2.761732 | 0.79 | 0.431 | -3.23808 | 7.587714 |
|  |  |  |  |  |  |  |
| **Constant** | 1.474746 | 1.742193 | 0.85 | 0.397 | -1.93989 | 4.88938 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 1.809977 | 1.022269 | 0.598297 | 5.475565 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 1.774071 | 1.100912 | 0.525717 | 5.986737 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 20.33417 | 0.702829 | 19.00227 | 21.75942 |
| **Academic system** | 5.733436 | 1.915343 | 2.978914 | 11.03499 |
| **Industry** | 3.270703 | 0.934917 | 1.867792 | 5.727348 |
| **Government** | 5.072872 | 1.117939 | 3.29359 | 7.813366 |
| **Health** | 8.503419 | 0.571205 | 7.454445 | 9.700002 |
| **Research Institute** | 13.7129 | 1.856314 | 10.51725 | 17.87955 |
| Number of observations = 639Log pseudolikelihood = -2678.2335Wald (10) = 458.98Log likelihood = -2678.2335 |

**Table S4**. Maximum-likelihood estimates from three-level random-intercept linear models of normalized citations with efficiency as covariate.

**S.5.2 Estimates from mixed-effects regression models of innovation**

Table S5 reports the full range of estimates from the hierarchical three-level random-intercept model of innovation including the local clustering coefficient as a covariate.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 6.42E-05 | 0.000109 | 0.59 | 0.555 | -0.00015 | 0.000278 |
| **Clustering coefficient (*cc*)** | -0.08197 | 0.028361 | -2.89 | 0.004 | -0.13756 | -0.02639 |
| **Closeness centrality (*l*)** | 0.21662 | 0.033234 | 6.52 | 0 | 0.151481 | 0.281758 |
| **Geographical entropy ()** | 0.162875 | 0.036848 | 4.42 | 0 | 0.090654 | 0.235095 |
| **Industrial collaboration (*ic*)** | 0.196698 | 0.06865 | 2.87 | 0.004 | 0.062146 | 0.33125 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | 1.060987 | 1.222795 | 0.87 | 0.386 | -1.33565 | 3.457621 |
| **Industry** | -0.37585 | 0.077794 | -4.83 | 0 | -0.52833 | -0.22338 |
| **Government** | -0.38585 | 0.07916 | -4.87 | 0 | -0.541 | -0.2307 |
| **Health** | -0.19238 | 0.079814 | -2.41 | 0.016 | -0.34881 | -0.03594 |
| **Research institute** | -0.09867 | 0.139943 | -0.71 | 0.481 | -0.37296 | 0.175608 |
|  |  |  |  |  |  |  |
| **Constant** | 0.32906 | 0.083979 | 3.92 | 0 | 0.164464 | 0.493656 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 5.95E-09 | 5.35E-08 | 1.30E-16 | 0.272217 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 2.97E-09 | 1.08E-08 | 2.39E-12 | 3.69E-06 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 1.357163 | 0.046495 | 1.269027 | 1.45142 |
| **Academic system** | 2.729741 | 0.865797 | 1.466034 | 5.082749 |
| **Industry** | 0.105642 | 0.028544 | 0.062208 | 0.179403 |
| **Government** | 0.142781 | 0.035741 | 0.087417 | 0.233209 |
| **Health** | 0.519942 | 0.033448 | 0.45835 | 0.58981 |
| **Research Institute** | 0.665177 | 0.090613 | 0.50931 | 0.868743 |
| Number of observations = 639Log pseudolikelihood = -914.73492 Wald (10) = 181.68Log likelihood = -914.73492  |

**Table S5**. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with the local clustering coefficient as covariate.

As clustering increases by one standard deviation, an organization’s innovation value decreases by 0.0820 units. An increase by one standard deviation in (normalized) closeness centrality was associated with an expected increase of 0.2166units in innovation value. After controlling for the other covariates, a one-unit increase in geographical entropy was associated with an expected increase in innovation value by 0.1629 units. Estimates suggested that there was no significant difference in innovation value between two organizations that differed in number of publications, after controlling for the other covariates.

Table S6 reports the full range of estimates from the hierarchical three-level random-intercept model of innovation in which the local clustering coefficient has been replaced by efficiency. Results are in qualitative agreement with the ones reported in the main text. A surge in opportunities of knowledge brokerage are associated with an expected increase in innovation value, after controlling for the other covariates. Moreover, unlike what was found in the case of research impact, the association between brokerage and innovation is now statistically significant.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 7.05E-05 | 0.000113 | 0.62 | 0.534 | -0.00015 | 0.000293 |
| **Efficiency (*eff*)** | 0.095779 | 0.029677 | 3.23 | 0.001 | 0.037613 | 0.153945 |
| **Closeness centrality (*l*)** | 0.201033 | 0.037094 | 5.42 | 0 | 0.128331 | 0.273736 |
| **Geographical entropy ()** | 0.22307 | 0.041761 | 5.34 | 0 | 0.14122 | 0.304921 |
| **Industrial collaboration (*ic*)** | 0.215189 | 0.06947 | 3.1 | 0.002 | 0.07903 | 0.351347 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | 1.077486 | 1.215413 | 0.89 | 0.375 | -1.30468 | 3.459652 |
| **Industry** | -0.29626 | 0.082758 | -3.58 | 0 | -0.45846 | -0.13406 |
| **Government** | -0.39026 | 0.086106 | -4.53 | 0 | -0.55903 | -0.2215 |
| **Health** | -0.16464 | 0.080795 | -2.04 | 0.042 | -0.32299 | -0.00628 |
| **Research institute** | -0.06094 | 0.137587 | -0.44 | 0.658 | -0.33061 | 0.208727 |
|  |  |  |  |  |  |  |
| **Constant** | 0.275586 | 0.092017 | 2.99 | 0.003 | 0.095237 | 0.455935 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 3.97E-11 | 4.13E-10 | 5.66E-20 | 0.027876 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 0.150211 | 0.063125 | 0.065916 | 0.342303 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 1.330701 | 0.046708 | 1.242232 | 1.425469 |
| **Academic system** | 2.712254 | 0.858253 | 1.458747 | 5.042903 |
| **Industry** | 0.098093 | 0.034607 | 0.049129 | 0.195857 |
| **Government** | 0.146177 | 0.040164 | 0.08531 | 0.250469 |
| **Health** | 0.515374 | 0.034133 | 0.452634 | 0.58681 |
| **Research Institute** | 0.643473 | 0.087905 | 0.492319 | 0.841033 |
| Number of observations = 639Log pseudolikelihood = -912.65477 Wald (10) = 174.24Log likelihood = -912.65477  |

**Table S6**. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with the efficiency as covariate.

**S.5.3 Estimates from mixed-effects regression models with interaction effects**

To check for robustness, we further estimated a number of additional three-level random-intercept models. In these models we introduced a level-1 interaction term between (standardized) geographical entropy and the clustering coefficient. Tables S7 and S8 report estimates from such models of research impact and innovation, respectively. Once again, results (and especially the main effects of the interacted covariates) remained qualitatively similar to the ones reported in the main manuscript. In addition, the parameter estimate for the interaction term was found to be negative and statistically significant in both models. That is, the expected negative effects on an organization’s research impact and innovation of an increase by one standard deviation in clustering are amplified as the geographical diversity of an organization’s collaboration network also increases by one standard deviation, after controlling for the other covariates. In other words, these results suggest that the benefits of open structures can be further amplified when an organization’s collaborators are spatially distributed across many and different countries.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 0.002789 | 0.002043 | 1.37 | 0.172 | -0.00121 | 0.006793 |
| **Clustering coefficient (*cc*)** | -0.42642 | 0.574581 | -0.74 | 0.458 | -1.55258 | 0.699739 |
| **Closeness centrality (*l*)** | 4.808166 | 0.723418 | 6.65 | 0 | 3.390292 | 6.22604 |
| **Geographical entropy ()** | 4.956622 | 0.419531 | 11.81 | 0 | 4.134355 | 5.778888 |
| **Industrial collaboration (*ic*)** | 4.570066 | 1.163106 | 3.93 | 0 | 2.290421 | 6.849712 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | 0.824126 | 2.032662 | 0.41 | 0.685 | -3.15982 | 4.808071 |
| **Industry** | -3.27789 | 1.383478 | -2.37 | 0.018 | -5.98946 | -0.56633 |
| **Government** | -6.28619 | 1.85707 | -3.39 | 0.001 | -9.92598 | -2.6464 |
| **Health** | -1.24359 | 1.273007 | -0.98 | 0.329 | -3.73864 | 1.251457 |
| **Research institute** | 2.466972 | 2.466364 | 1 | 0.317 | -2.36701 | 7.300957 |
|  |  |  |  |  |  |  |
| **Constant** | 2.966713 | 1.680229 | 1.77 | 0.077 | -0.32648 | 6.259902 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 1.69721 | 1.125712 | 0.462549 | 6.227501 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 2.409696 | 0.939308 | 1.122434 | 5.173253 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 19.66172 | 0.676594 | 18.37936 | 21.03357 |
| **Academic system** | 3.627777 | 1.421989 | 1.682647 | 7.821469 |
| **Industry** | 1.879379 | 0.539752 | 1.070414 | 3.299718 |
| **Government** | 4.666399 | 1.045263 | 3.008255 | 7.238511 |
| **Health** | 8.768805 | 0.617094 | 7.639028 | 10.06567 |
| **Research Institute** | 12.06206 | 1.66891 | 9.197057 | 15.81954 |
| Number of observations = 639Log pseudolikelihood = -2658.2566 Wald (10) = 549.88Log likelihood = -2658.2566  |

**Table S7**. Maximum-likelihood estimates from three-level random-intercept linear models of research impact with interaction effects.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Coef.** | **Std. err.** | **z** | **P>|z|** | **[95% Conf. interval]** |
|  |  |  |  |  |  |  |
| **Number of articles in WOS (*v*)** | 0.000132 | 0.000104 | 1.27 | 0.203 | -7.2E-05 | 0.000336 |
| **Clustering coefficient (*cc*)** | -0.06273 | 0.029169 | -2.15 | 0.032 | -0.1199 | -0.00556 |
| **Closeness centrality (*l*)** | 0.165457 | 0.034408 | 4.81 | 0 | 0.098018 | 0.232896 |
| **Geographical entropy ()** | 0.075082 | 0.021854 | 3.44 | 0.001 | 0.032249 | 0.117914 |
| **Industrial collaboration (*ic*)** | 0.232026 | 0.063213 | 3.67 | 0 | 0.108131 | 0.355921 |
|  |  |  |  |  |  |  |
| **Type of organization** |  |  |  |  |  |  |
| **Academic system** | 1.178321 | 1.185711 | 0.99 | 0.32 | -1.14563 | 3.502272 |
| **Industry** | -0.21548 | 0.080932 | -2.66 | 0.008 | -0.3741 | -0.05686 |
| **Government** | -0.31996 | 0.085176 | -3.76 | 0 | -0.4869 | -0.15302 |
| **Health** | -0.13869 | 0.081155 | -1.71 | 0.087 | -0.29775 | 0.020375 |
| **Research institute** | -0.05406 | 0.126833 | -0.43 | 0.67 | -0.30265 | 0.194529 |
|  |  |  |  |  |  |  |
| **Constant** | 0.290886 | 0.088449 | 3.29 | 0.001 | 0.11753 | 0.464242 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Random-effects parameters** | Estimate | Std. Err. | [95% Conf. Interval] |
|  |  |  |  |  |
|  |  |  |  |  |
| **Geographical region )** | 2.58E-09 | 2.60E-08 | 6.54E-18 | 1.014793 |
|  |  |  |  |  |
|  |  |  |  |  |
| **Country )** | 0.166526 | 0.056729 | 0.085411 | 0.324676 |
|  |  |  |  |  |
| **Level-1 residual by type of organization ()** |  |  |  |  |
| **Academic** | 1.308692 | 0.04497 | 1.223455 | 1.399866 |
| **Academic system** | 2.645774 | 0.840066 | 1.419992 | 4.929691 |
| **Industry** | 0.076094 | 0.025267 | 0.039693 | 0.145878 |
| **Government** | 0.118277 | 0.034155 | 0.067159 | 0.208306 |
| **Health** | 0.534018 | 0.036643 | 0.46682 | 0.610891 |
| **Research Institute** | 0.58007 | 0.080519 | 0.441902 | 0.761438 |
| Number of observations = 639Log pseudolikelihood = -903.79064 Wald (10) = 203.47Log likelihood = -903.79064  |

**Table S8**. Maximum-likelihood estimates from three-level random-intercept linear models of innovation with interaction effects.

# **References**

1. Newman MEJ. *Networks. An Introduction.* Oxford, UK: Oxford University Press; 2010.

2. Newman MEJ. Scientific collaboration networks. II. Shortest paths, weighted networks, and centrality. *Physical Review E*. 2001;64(1):016132.

3. Garas G, Cingolani I, Panzarasa P, et al. Network analysis of surgical innovation: Measuring value and the virality of diffusion in robotic surgery. *PLoS One.* 2017;12:e0183332.

4. Opsahl T, Panzarasa P. Clustering in weighted networks. *Social Networks*. 2009;31:155-63.

5. Latora V, Nicosia V, Panzarasa P. Social cohesion, structural holes, and a tale of two measures. *Journal of Statistical Physics*. 2013;151:745-64.

6. Burt RS. *Structural Holes. The Social Structure of Competition*. Cambridge, MA: Harvard University Press; 1992.

7. Opsahl T, Agneessens F, Skvoretz J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Social Networks*. 2010;32:245-51.

8. Savage N. Industry links boost research output. *Nature.* 2017;552:S11-S13.

9. Skrondal A, Rabe-Hesketh S. *Generalized Latent Variable Modeling. Multilevel, Longitudinal, and Structural Equation Models*. Boca Raton, FL: Chapman & Hall/CRC; 2004.