# SCIENTIFIC REPORTS

**OPEN**

# Driving forces of researchers mobility

Floriana Gargiulo & Timoteo Carletti

NaXys, University of Namur, Namur, Belgium.

Correspondence and
requests for materials
should be addressed to
F.G. (floriana.
gargiulo@gmail.com)

Starting from the dataset of the publication corpus of the APS during the period 1955–2009, we reconstruct the individual researchers trajectories, namely the list of the consecutive affiliations for each scholar. Crossing this information with different geographic datasets we embed these trajectories in a spatial framework. Using methods from network theory and complex systems analysis we characterise these patterns in terms of topological network properties and we analyse the dependence of an academic path across different dimensions: the distance between two subsequent positions, the relative importance of the institutions (in terms of number of publications) and some socio–cultural traits. We show that distance is not always a good predictor for the next affiliation while other factors like "the previous steps" of the career of the researchers (in particular the first position) or the linguistic and historical similarity between two countries can have an important impact. Finally we show that the dataset exhibit a memory effect, hence the fate of a career strongly depends from the first two affiliations.

Recently a large scientific interest has grown around the analysis of scientific production. In this framework many recent papers have been published concerning "scientometrics", namely the methods to evaluate the scientific production and to forecast the possible success of scientists[3–6]. This is the reason why, in such papers, the focus has been put on the number of published papers in each individual career.

The idea of using scientific publications in the framework of the complex network science is not new, several studies have been proposed[7–9,24,25] where the structure and the topological properties of collaborations and citations networks were analysed. In these cases the attention was mainly focused on the citations patterns, the coauthorship communities and on the mechanisms leading to the success of papers in terms of citations. A recent interesting work in this framework is[10] where the importance of spatial features in collaborations and citations networks was analysed. The authors showed that geography has a leading role in the morphogenesis of these networks and that the strength of the relationships can be modelled by a gravity law.

In the present paper, still dealing with the scientific production data, we focused our research on the mobility in the framework of the research job market. Our aim is to understand which are the driving forces responsible for the choice of an academic position. Some studies on this topic have been developed by the same institutions providing fellowships for international careers (like the Marie Curie) or in the framework of sociological analysis concerning the *talents* migration paths and their impact on social inequalities[11–13]. In all these cases the analysis was based on surveys, concerning a sample that cannot always be considered to be significant. Moreover, most of these studies were focused to a precise geographical area (mainly at country level), analysing mostly the impact of the local recruitment policies on the academic careers.

We use a complex systems approach to analyse the researchers mobility paths, reconstructing a large number of careers starting from a large corpus of papers published on the journals of the American Physical Society, the spatial trajectory of each author being identified looking at the sequence of its affiliations and their spatial embedding; we hereby assume two subsequent affiliations to be a proxy for the researcher mobility. Let us observe that, while our dataset has the evident limit to be addressed mostly to the physics community, we have on the other hand at our disposal an extremely large number of trajectories to analyse and thus be able to rely our results on good statistics.

We first focus on the individual behaviours by studying the researchers' trajectories at two levels of granularity: the paths through the universities and the paths through the countries. In this framework a recent literature has been developed concerning human mobility[14–17]. Even if in the present case we can't apply the same statistical tools, given the intrinsic lower length of the researchers paths, we can use a similar formalism concerning the career reconstruction: we study both the trajectories formed by legs regardless of the time duration of each shift between universities and also the trajectories where to each leg, the time duration and the beginning of the affiliation period is associated to, therefore the time dimension is fully taken into account in this second case.

Secondly we consider an aggregate scale, focusing on the roles of universities and countries, using a static network approach[20,21]. We reconstruct the bipartite graph authors - universities (or countries) and its

mono–partite projections: the networks of universities and of countries. We analyse the basic topological properties of these graphs and then we study the characteristics of their embedding in the geographical space. Finally we will conclude by studying the impact of the university rank, in terms of publications, on the paths, analysing both the single transitions and the processes with memory.

## The dataset

The data we use are extracted from the American Physical Society[2] publications database containing the records of all the paper issued in the APS journals from 1955 to 2009. Each paper represents one point of the careers of the authors. In such a way, for $N_R = 71.246$ researchers in the database, we can reconstruct a trajectory constituted by the temporal sequence of all the distinct affiliations of their papers. We should notice that the considered careers are derived by a sample of all the authors' publications, namely the ones issued on the APS journals, therefore our data could miss some steps in the careers reconstruction of researchers mobility. We are however confident in our results because the APS covers a very large subset of all the publications in the physics domains. We also observe that the number of countries and universities present in the dataset evolves in time, mostly due to historical and societal changes. In the very first issues of the journals only US and few western European countries are represented, while in the latest decades these journals become a worldwide reference for the physical society (see Figure 1 in the SI). The trajectories are embedded in the geographical space using the information extracted by the CEPII geographical database[1] (GIS coordinates for each city, distances between countries, historical and linguistic correlations). More information on the data preprocessing can be found in the SI.

All the individual paths can be aggregated in time and then projected to give rise to two possible directed weighted networks (Figure 1): the universities networks (where the nodes are the $N_{univ}$ = 2434 universities) and the countries network (the nodes are the $N_{countries}$ = 174 countries). In both cases two nodes are connected if at least one researcher moves among them (the weight, $w$, of a link is the number of researchers moving among the two nodes).

## Results

**Individual level: the researchers paths.** The length $l$ of a path is given by the total number of its legs, even if the path passes more than once through the same university; this information has not been lost, being captured by studying the total number of distinct countries and universities each researcher has visited, hereby denoted respectively by $n_{countries}$ and $n_{univ}$. We first observe that a large part (more than 40%) of the paths have length $l = 1$, namely the most probable careers are composed by two academic positions and a single movement among these. This is also confirmed by the distribution of the number of visited universities; on the other side, if we focus on the number of visited countries, we can observe that an important portion of paths just remain in a single state (see the Figure 2 in the SI). In general we can notice that the 90% of researchers visit at most 4 universities and 3 countries. These results are stable considering careers starting in different time periods (see the Figure 3 in the SI) showing that researchers paths follows prototypical rules that are not evolving in time (PhD+1 postDoc+Permanent position, PhD+ Tenure Track, PhD+Postdoc+Leave Academy…).
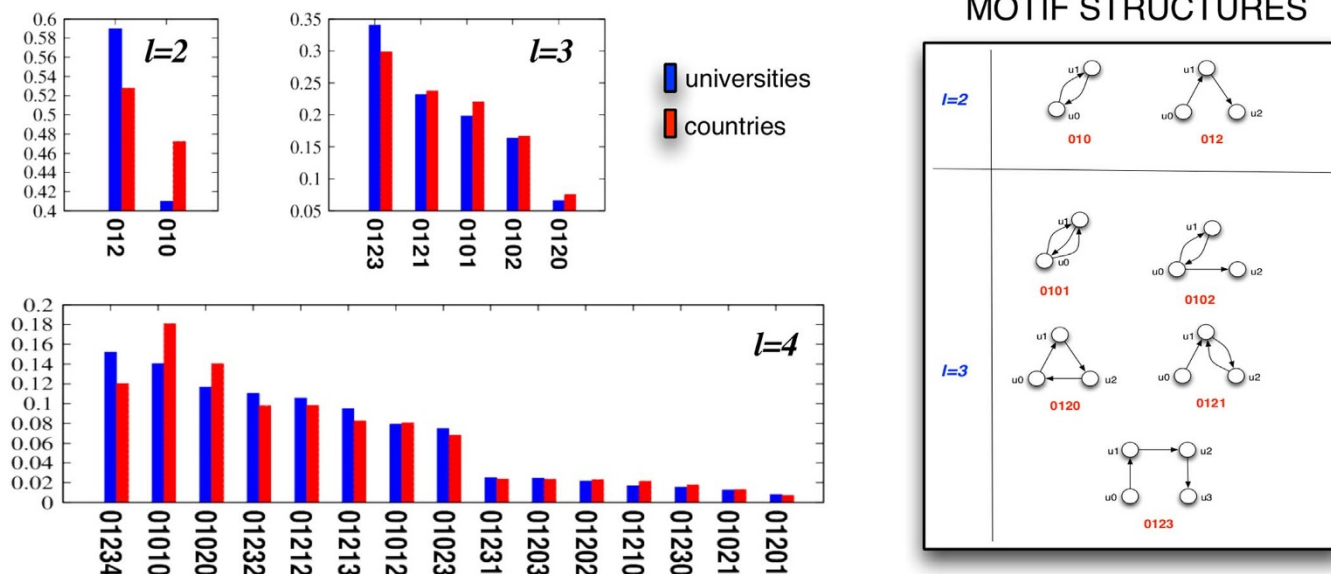
In Figure 2 we analyse the topological structures, i.e. the exhibited motifs, of the paths (focusing on the paths of length 2, 3 and 4). We can observe that for low length paths ($l = 2, 3$) most of the trajectories connect different universities, paths 012 and 0123. This tendency is a bit less marked at the level of states, showing a certain tendency to have more than one position in the same state, paths 010 and 0121 or 0101). For longer paths, say $l = 3$ and $l = 4$, we observe that many patterns are just between two universities, path 01010. Finally we can notice a very high tendency to return, at some point of the career, to
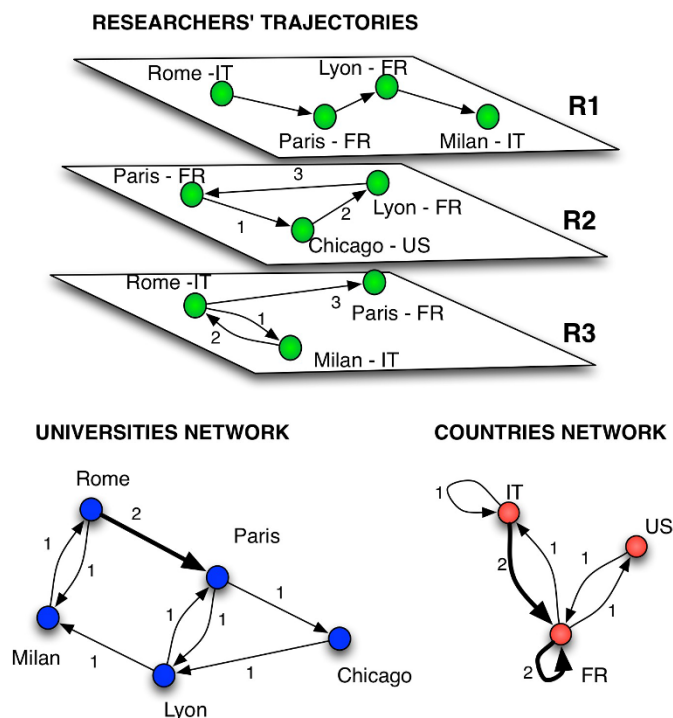


**Figure 1 | Researchers trajectories, time aggregate universities network and countries network.** Top panel: three prototypical researchers trajectories of length three. Each node on a path represents an university the researcher has visited during her/his career, a directed link is drawn between two nodes whenever the researcher moved from one university to a second one, the arrow pointing from the past affiliation to the next one, to denote the time flow direction. R1: four different universities have been visited, R2: the researcher ended her/his career in the same university she/he started visiting thus 3 universities, R3: after a stay in Milan, the researcher went back to her/his career origin and then moved away. Bottom left panel: the universities network. Nodes denote universities and a directed weighted link is drawn whenever a researcher moved from one university to a second one, the arrow denotes thus the net flow of scholars from one university to another one during the period 1955–2009, the weight being the total number of such moves. Bottom right: the countries network. Nodes represents countries, a directed weighted link between two countries denotes the total number of scholars leaving any university belonging to the first country and going to any university of the second one during the time period under scrutiny.

the origin point where the career started, probably corresponding to the place where the personal life of the researcher is centred. These findings have been checked against a null–model. We created a reshuffling algorithm allowing us to preserve the original path lengths distribution and the in and out degrees for each university or country (see SI for a more detailed description). Once applied on our dataset, the new obtained networks will not show anymore such particular motifs distribution, that thus results to be peculiar to the original data. This fact can be quantified computing the metric entropy of each path, both for the real network and for the reshuffled model; even if the most entropic careers (012, 0123, 01234) are largely present in the real dataset, their frequency is strongly overestimated by the null model. At the same time the reshuffled model is not able at all to reproduce the minimal entropy paths (i.e. the ones connecting just two nodes). The comparison with the base model therefore gives an indication of the fact that the less entropic paths are a typical characteristic of the system (see Figure 5 in the SI).

The importance of the first position in the career can be also observed in Figure 3 where the average of all the trajectories considered as sequences of legs regardless of the time duration, shows a

**Figure 2 | Motifs distribution in the universities and countries networks.** Limiting the presentation to paths of length $l = 2$, 3 and 4 we show the frequency of the topological structures (linear trees, round trips, 3–cliques, etc.) present in the above networks. The motif structures for $l = 2$, 3 are displayed in the right panel. From our data we can conclude that researchers with short careers visit mainly different universities, while they tend to remain in the same country showing a propensity to have more than one position in the same state. In the case of longer careers we remark a strong tendency to move back and forth between two universities or two countries, fact that can be associated with a double affiliation of the researchers more than to real mobility.

gradual spread from the origin with a marked alternated tendency to get closer to the starting point. The lower panel of Figure 3 concerns the trajectories where the time dimension has been taken into account ($t = 0$ corresponds to the first paper published with an affiliation, the steps duration correspond to the period of time between two consecutive publications with different affiliations). In this framework, considering couples of trajectories starting from the same position (each one with its own time origin $t = 0$) we can analyse their relative distance each year. We report two characteristic behaviours, Paris (Fr) and Berkley (US), and in both cases we observe that after a fast separation during the first 6–7 years a stable point is roughly obtained. Let us remark that a similar behaviour has been obtained for all the largest universities and thus it is robust across careers in our database. This suggests that, on average, the time before the stabilisation of the careers is approximately 6.39 years (see SI for more details). Let us emphasise that our result is in good agreement with the one obtained in[26] analysing different disciplines and applying completely different metrics. We nevertheless remark two different asymptotic behaviours; careers started at Berkley still continue to diffuse, even if slowly, showing that the US university acts as a career spreader, on the other hand in the case of Paris we observe a tendency to get closer to the career origin in the late years of the academic life.

**The researchers mobility network.** We start analysing the standard topological quantities of the network structure, both for countries and universities. In the case of countries network self-loops, namely the movements between two universities in the same countries, have an important role and therefore have been considered separately. The self-loops are not taken into account for the calculation of the strengths of the nodes ($s_{in(out)}^i = \sum_{j \neq i} w_{ji(ij)}$, $s_{tot}^i = \sum_{j \neq i} (w_{ji} + w_{ij})$).

As we can observe in Figure 4A both for countries and universities, the weights follows a power law distribution, demonstrating a very different connectivity behaviour among couple of nodes. For universities and countries, excluding self–loops the slope of the frequency curve is the same ($P(w) = w^{-2.1}$) while for the self–loops the slope is lower $(P(w_S) = w_S^{-1.1})$. The role of self–loops for each node can be estimated with the *endogamy* index: $\epsilon_i = w_{ii}/(s_{tot}^i + w_{ii})$. As we can observe this index has two different behaviours with respect to the total strength (Figure 4B): it first decreases with $s_{tot}$ and then after a certain value it shows the opposite tendency. This is due to the fact that some countries, notably the not well connected ones or the not much active ones, have no sufficient international impact to exchange researchers with other countries, and therefore the internal (or at very local scale) mobility is favoured. Of course as the strength increases the international exchanges increase (going outside the country is necessary to increase the personal researchers' impacts) and the endogamy decreases. On the other side, reached a certain level of importance of the countries, researchers could be interested, for linguistic or social reasons, to remain in the same country. The larger is the strength of the country the lower is the effect of the choice of the local mobility on the researchers impacts. This explains the tendency inversion, after a certain value of $s_{tot}$, observed in the figure.

Figure 4C shows the strength of the first neighbours of a node $\left(s_{nn}^i = \sum_{j \in v(i)} s_{tot}^j\right)$ as a function of its total strength. Both for countries and universities there is a weak dissortative tendency (lower strength nodes tend, on average, to connect to higher degree nodes). On the other side, observing the scatter plot of the single measures, we can notice that low strength nodes are also frequently connected in low strength communities.

The clustering coefficient (Figure 4D) is the classical measure of the level of cohesiveness of the neighbourhood of a node[20], $c_i = \frac{1}{2k_i(k_i-1)} \sum_{j,k \neq i} a_{ij}a_{ik}a_{jk}$, where $a_{ij}$ are the elements of the adjacency matrix of the unweighted network. If $c_i = 0$ the node is in a tree like structure where the neighbours are not connected at all, if $c_i = 1$, it is in a completely connected environment. The weighted clustering coefficient is defined as, $c_i^w = \frac{1}{2s_i(k_i-1)} \sum_{j,k \neq i} (w_{ij} + w_{ik}) (a_{ij}a_{ik}a_{jk})$.
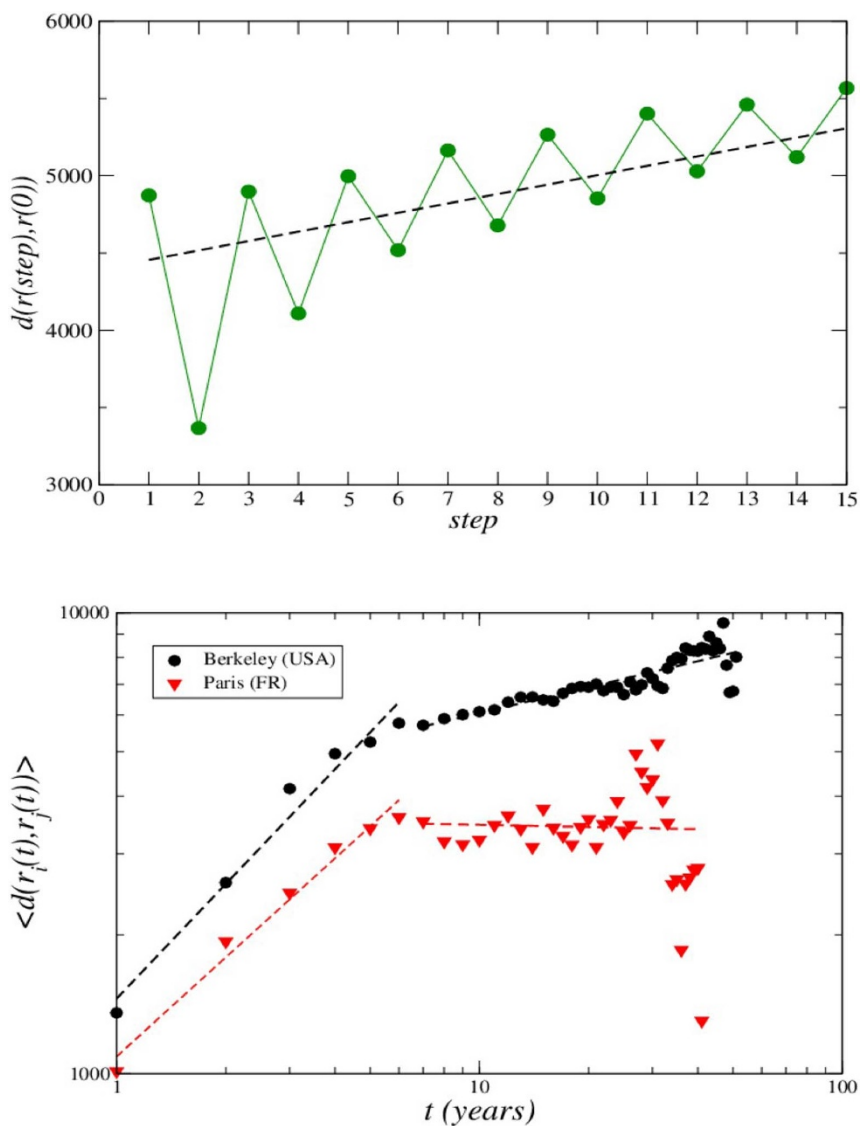
**Figure 3 | Spatial properties of the average trajectory.** We compute the spatial distance between two consecutive affiliations and we studied it as a function of the number of different affiliation and also as a function of the duration of the career measured in years. Upper plot: average distance from the career origin of the successive affiliations, where each trajectory is formed by subsequent steps, regardless of the time duration. The average is calculated for all the trajectories in the dataset. We can clearly see the combination of two antagonistic processes, a diffusive one, tending to push researcher to visit farther and farther universities, coupled with a tendency to get closer to the career origin, probably where the personal life of the researcher is centred. Lower plot: average distance for any couple of trajectories starting from the same career origin as a function of the career duration. For two origins, Paris (Fr) and Berkley (US), we can observe a well defined trend, that we found to be generic across our dataset: initially scholars' trajectories tend to separate far apart quite fast, while in a second phase this behaviour is reduced but still at work (case of Berkley) or even inverted (case of Paris) showing again a tendency to get closer to the career origin in the late phase of the researcher career.

If $c_i^w > c_i$ then the tendency to form triangles is influenced by the weights structures and not only by the topology[21]. For the universities network the clustering coefficient increases with the total strength of the nodes, showing a higher tendency for stronger institutions to be inside a fully connected scientific community. The weighted clustering is always higher than the unweighted one, demonstrating that the cliques formation is determined by the weights of the links. For the network of countries we observe again a high clustering for low connected nodes, that is another indicator of the presence of local low degree cliques.

In Figure 5 we analyse the correlation between the network morphology and some geo-cultural traits such as the distances, the language and the historical links between the involved countries and universities. In spatial networks[22] like mobile phone[15,16] or commut-

ing networks[18,19], a dependency between the weights of the links and the distance is observed, usually reproducible through a gravity law. This is due to the fact that the choice of the destination is mostly determined by the traveling distance between the points. For researchers, the mobility choices are also influenced by the heterogeneity of the *importance* of the institutions and by the existence of collaboration communities, therefore the direct correlation with the geographic space is partially lost. To explore such phenomenon, we introduce a quantitative indication of the scientific impact of each university, by dividing them into 5 groups according to their number of publications in the period 1955–2009: (R1) less than 50 papers, (R2) between 50 and 100 papers, (R3) 100 and 1000 publications, (R4) 1000 and 10000 and (R5) more than 10000 papers. For the links starting from low rank nodes we can observe that the correlation with
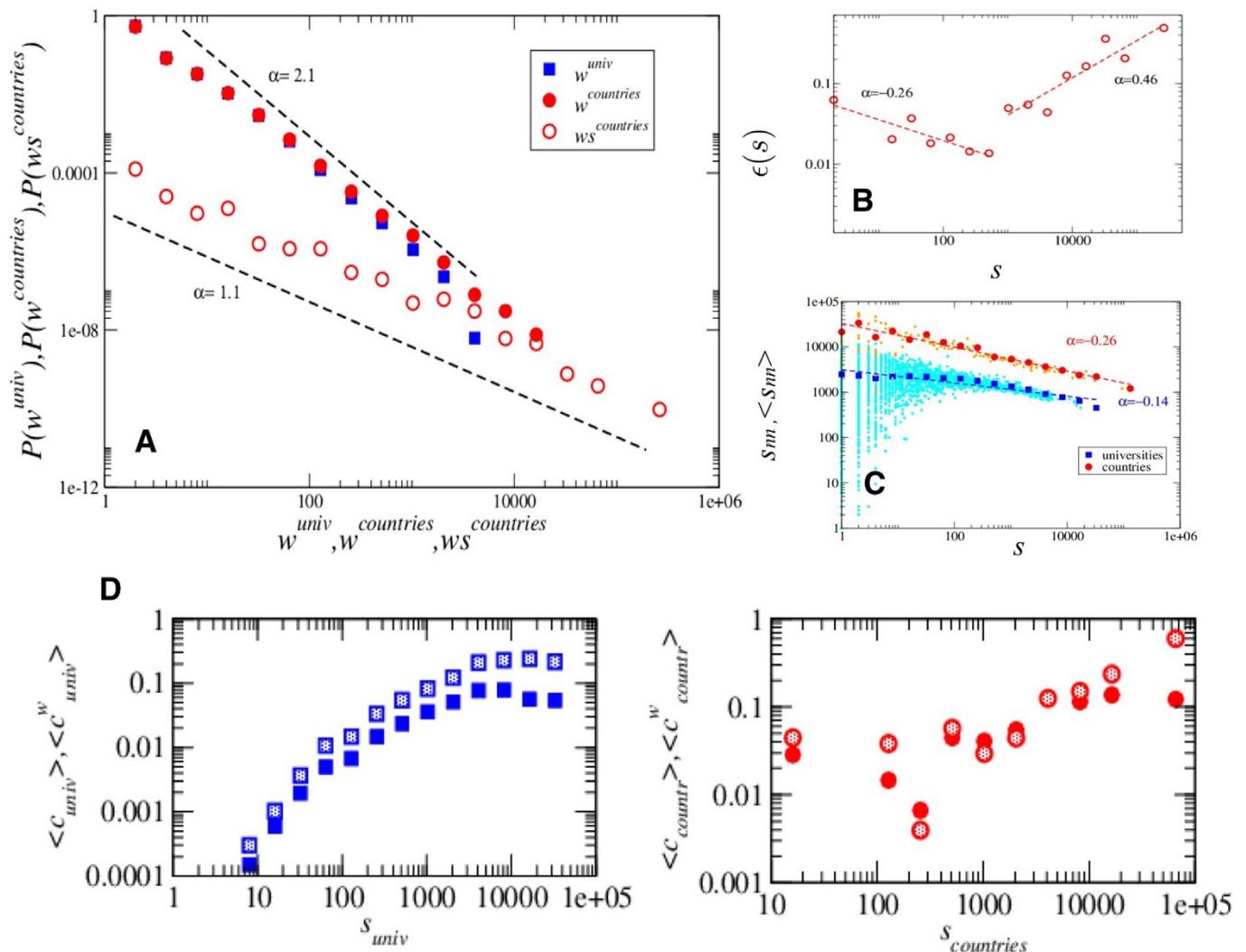
**Figure 4 | Universities and countries weighted networks.** Plot A: Weights distribution for the university network (blue squares), weights distribution for the countries network excluding self–loops (red full circles) and distribution of self–loops weights for countries (red empty circles). All the distributions are broad and display a scale free behaviour. Plot B: Fraction of self–loops with respect to the total number of connections of the nodes as a function of the strengths for the country network. Using the endogamy index, $\epsilon_i = w_{ii}/\left(s^i_{tot} + w_{ii}\right)$, we can enlighten a non monotone dependence with respect to the total strength, for countries where the international exchanges are very limited there are mainly endogenous movements inducing a decreasing of endogamy as long as few international exchanges are realised; there exists however a tipping point beyond which the more international exchanges are present the more the country acquire international prestige and thus inner scholars have an incentive to rest in the same country, explaining thus the tendency of $\epsilon_i$ to increase with $s_{tot}$. Plot C: Average total strength of the neighbours as a function of the total strength of the nodes; orange and red circles represent respectively the scattered and averaged values for countries network and cyan and blue squares for universities. One can observe a weak tendency of lower strength nodes to connect to higher degree nodes, that is a dissortative behaviour; however the scatter plot of the individual data shows that low strength nodes are also frequently connected in low strength communities. Plot D left panel: Clustering coefficient (full squares) and weighted clustering coefficient (empty squares) as a function of the total strength of the nodes for the universities network. Plot D right panel: Clustering coefficient (full circles) and weighted clustering coefficient (empty circles) as a function of the total strength of the nodes for the countries network. Because the weighted clustering coefficient is larger than the unweighted one, both for universities and countries, we can conclude that exchanges among the universities or countries are sustained by the weights structures and not only by the topology. Moreover the clustering coefficients for the universities network, increase with total strength of the nodes, showing the tendency of strong universities to be part of well connected collaboration networks, even if less marked this trend is also present in the case of countries network.

geography is maintained: large weights correspond to low distances and viceversa. As the rank increases we can observe that this correlation persists only for extremely large weights.

Calculating the distance between states is a non-trivial task. Usual measures based on the distances between the capitals didn't allow us to find any correlation between the weights and the distances for the countries network. Therefore we considered the fraction of fluxes for each node pointing to a contiguous country, to a country speaking the same language or to a country connected by the historical traits. We found that only the 25% of the countries have a high fraction of

fluxes (larger than 30%) pointing to a node in one of the previous cases. Therefore we can conclude as a first result that the more obvious drivers for human migrations patterns (geographical and cultural contiguity) are not sufficient to characterise the researchers mobility. For the 25% of countries for which these factors are significant we studied which of the two dimensions (geographical or cultural) is the more important one. In the lower plot of Figure 5 the cultural dimension, language and historical traits, is represented on the vertical axis and the geographical one on the horizontal axis. In the plane we can identify clusters of countries with similar prefer-
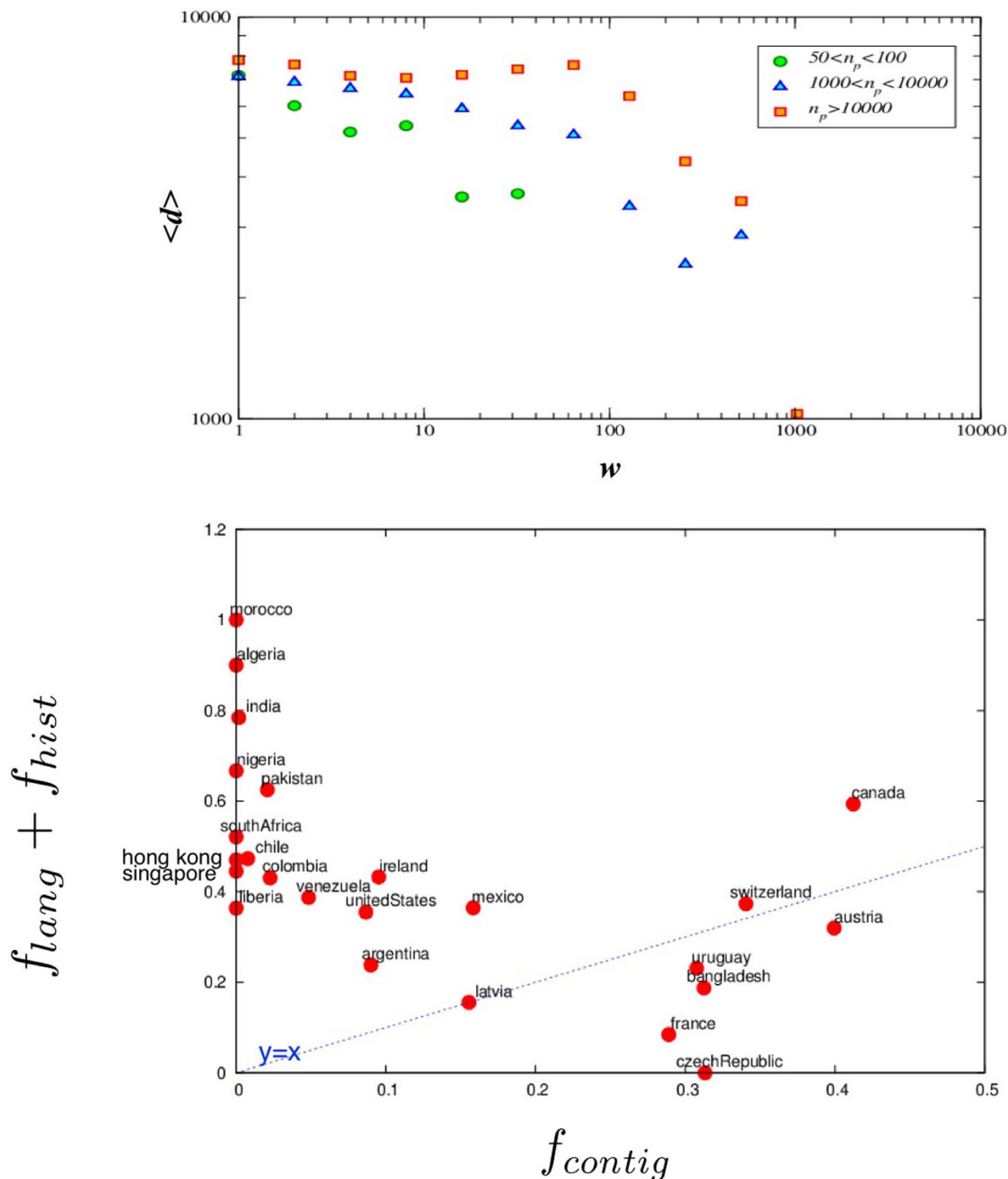
**Figure 5 | The importance of distance, language and culture.** Upper panel: Average distance ($km$) between two nodes as a function of the weights among them for the universities network; green circles represent jumps starting from a low rank university (number of publications: $50 < n_p < 100$), blue triangles from an intermediary rank ($1000 < n_p < 10000$), and red squares from the top rank ($n_p > 10000$). We can observe that scholars' displacements among low ranked universities exhibit correlations with geography similar to the ones presents also in other human movements scenarios. However as long as the rank increases such correlations tend to disappear, showing that distance is not a restraint once researchers move across well reputed universities. Correlations do persist, for high ranked universities, only for very large weights, where even the incentive to reach a good university could not be enough to induce a very large movement. Lower panel, spatial contiguity versus socio–cultural traits: Each point represents a country with at least 30% of its flows toward either a country with the same language or connected by an historical link either to a contiguous country. On the vertical axis we represent the fraction of flows to a country with the same language ($f_{lang}$) or connected by an historical link ($f_{hist}$). On the horizontal axis we report the fraction of flows to a contiguous country ($f_{contig}$). We can clearly identify two well defined clusters: countries for which the spatial proximity doesn't matter too much, while the exchanges are mainly due to common shared language and/or past history, e.g. Morocco and Algeria with France, India–Pakistan and South Africa with UK, Chile–Colombia and Venezuela with Spain. The second cluster is composed by countries whose flows are mainly directed to neighbours, this is the case of Canada due to its proximity to US, Switzerland and Austria because of their centrality in Europe, but also relatively small countries close to larger ones, e.g. Bangladesh or Uruguay respectively neighbours of India or Argentina and Brazil.

ences. For Canada, due to the border with the US, and for Switzerland and Austria (due to centrality in Europe and diffusion of their official languages) both the dimensions are equally important. For other central-European countries, e.g. France, and for small countries contiguous to an important one, e.g. Bangladesh or Uruguay, the geographical dimension is largely prevalent. On the other side we have several cases where the cultural dimension is prevalent, for instance the spanish speaking community, Chile–Colombia–Venezuela, the case of France with the ex–colonies, Algeria and Morocco, and UK with India, Pakistan and South Africa.

**The importance of being (starting) important.** In this section we study how the choice of a destination country/university depends on the scientific relevance of the country/university in terms of number of publications.

In Figure 6 A, B we report the conditional probability to end up in a university/country with $n_p^{dest}$ publications starting from an university or country with $n_p^{orig}$ publications and - Figure 6 C, D - the correlations between the number of publications in the origin country/university and the average number of publications in the destination country/university for each leg of the authors' paths. The conditional probability computed from our dataset is compared with a null hypothesis model that preserves the network topology but reshuffles randomly the number of publications among the nodes. The aim of this randomised model is to emphasise the contribution to the conditional probability arising from the network topology alone, that is a clear preference for the most abundant kind of university nodes (i.e. those with a low number of publications) confirmed by the high values of the conditional probability for small $n_p^{dest}$ irrespective of the size of $n_p^{orig}$, and the importance of the endogamous links for the countries as testified by the peaks of $P$ for $n_p^{dest} \sim n_p^{orig}$. On the contrary the real data always exhibit a preference toward the most publishing universities/countries clearly visible from the large values that $P$ assumes for large $n_p^{dest}$ for any values of $n_p^{orig}$. This tendency is less marked for universities with small $n_p^{orig}$ that show a relevant contribution to small $n_p^{dest}$ as well. This different tendency translates into a weak assortative correlation - as one can infer from the quite small slope obtained with a best linear fit - between the number of publications in the origin country/university and the average number of publications in the destination country/university: less important nodes (in terms of publications) are, on average, connected with less important nodes and viceversa.

Let us observe that one could have argued the existence of a sort of clustering between universities or countries, where high ranked universities exchange scholars among them while scholars from low ranked universities do mainly remain in the same community of low ranked ones. So we think that the absence of a strong correlation between the rank of the origin university/country and destination university/country is an interesting results. Figure 6 E, F presents the analysis of the correlations between the number of publications of the origin and the destination universities/countries for the European and the US mobility network, more precisely we selected a subset of legs in our database for which origins and destinations were in Europe or US. The assortative tendencies observed for the whole network are somehow different if we consider the mobility network for Europe and for US. At the level of universities, the assortativity is stronger for US than for Europe. Moreover, at the country level, in Europe we observe a counterintuitive dissortative behaviour. This is due to the fact that inside each European state a large heterogeneity exists at the level of academic institutions; therefore, an excellent university can be surrounded by several other low ranked institutions, lowering the average of the country level. This point suggests the fact that at the level of the individual destination choice two different mechanisms can be identified: either a destination institution is selected based on the importance of this institution itself either a country is selected on the basis of other criteria (distance, language, historical links, etc…).

To deepen the analysis on the career choices according to the scientific relevance of the academies, we resort to the previous division of the countries and the universities into different categories based on the number of published papers and we studied the transition probabilities between the different categories. At the level of countries (Figure 7A), we can observe that a large part of the paths legs points to a high ranked country. In fact, even if these nodes are less numerous, they have the highest number of passing trough traffic. Moreover a sort of homophily between states is observed in the destination choice: many trajectories of the paths connect countries in the same category (notice that the connections between the same country are not considered in this study).

At the level of universities (Figure 7B), we can observe that in general good institutions (R4) ($1000 < n_p < 10000$) are the most selected destinations from all the categories. Excellent institutions (R5) ($n_p > 10000$) are reached with a higher probability from higher ranked origins.

We also considered a memory (2 legs) process in analogy of what presented in[23]. Comparing two steps transitions computed from data and the normalised product of the single steps transitions, we can observe that the system exhibits large memory effects and that the probability of reaching a destination is strongly influenced by the first visited institution. Moreover, we can observe that low ranked institutions, that also have on average a very small traffic, are reached with a high probability only from researchers that twice stayed in low level institutions and never by researchers with a past career in better institutions. Also the contrary is true: two consecutive affiliations into high ranked universities will be almost never followed by a low and medium ranked one.

## Discussion

In summary, we analysed the researcher mobility paths from a data driven point of view using the dataset of all the papers published in the American Physics Society journals from 1955 to 2009. We used two different approaches: first we focused on the individual trajectories and then we studied the system at the level of aggregate mobility network structures. The analysis of individual trajectories shows a preference for short paths (between 2 and 4 steps) and that the 90% of the researchers, in their careers visit at most 4 universities and 3 countries. The first point of the path has a particular importance, and the trajectory has an high probability to pass through this point several times. Moreover a sort of memory effect of the starting points is observed when we analyse the time evolution of the average of the distance of all the trajectories from their starting point: an oscillating–like behaviour tending to get closer to the origin point is superposed to a general diffusive dynamics.

The network structures, both at country level and at university level, present heterogeneous weights, degrees and strengths and a general tendency to strength (and degree) assortativity, even if this phenomenon is less pronounced in Europe than in US. Higher degree nodes are usually part of more connected communities (higher clustering coefficient). The role of geography in the network morphogenesis appear to be relevant only for less important universities. At the level of countries, contiguity seems to be determinant only in particular cases of very geographically central countries, while strong drivers for connectivity are the language and the historical colonial connections.

In general a large part of the steps of the careers tend to lead the researchers to higher ranked institutions (that are the ones offering most of the positions). On the other side we can observe an important role of the first point of the path: starting from a lower rank institution, independently by the following step, lowers the probability of reaching a top rank academy and makes higher the probability to
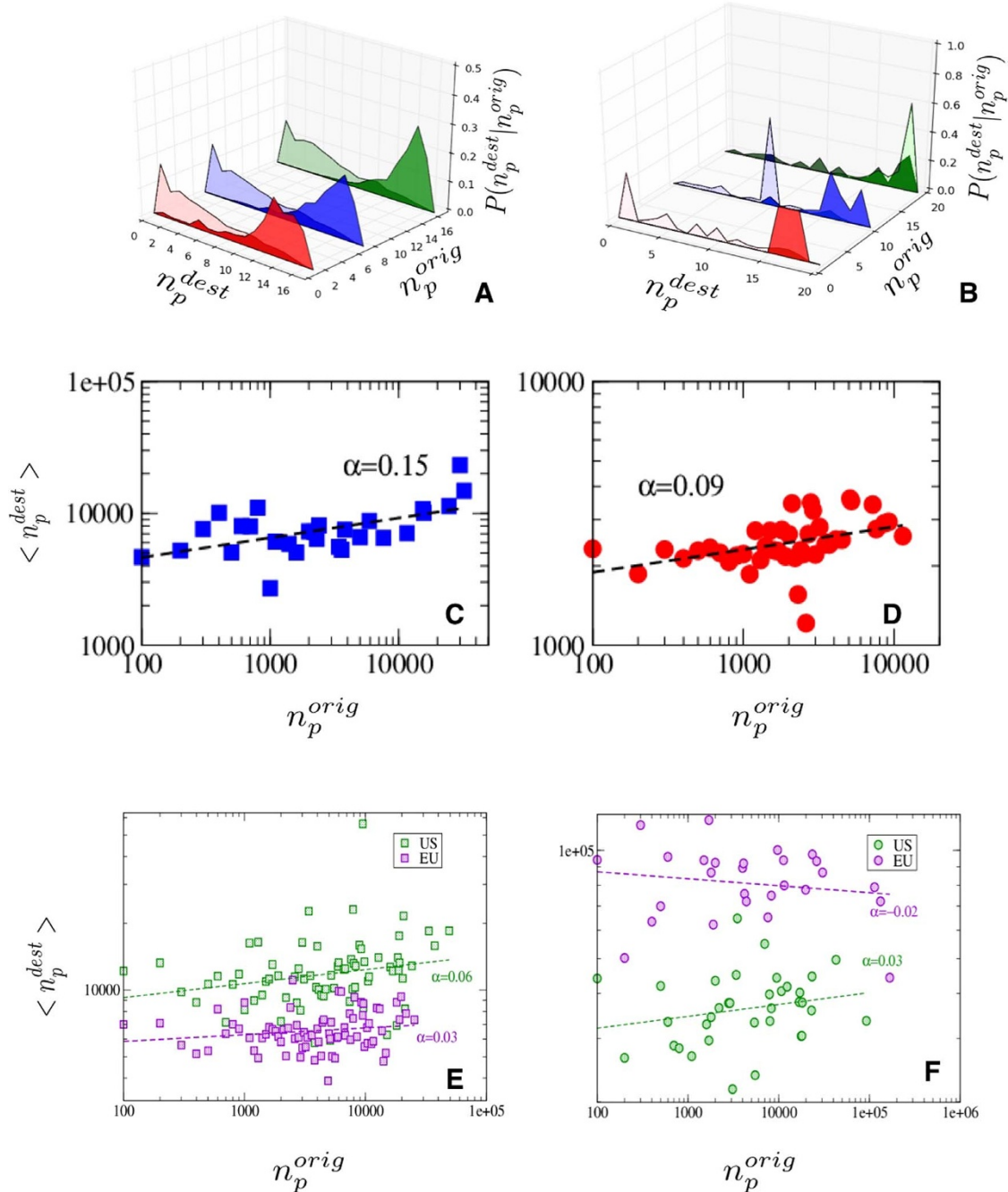
**Figure 6** | Panels A and B: Conditional probability to end up in a university (A) or country (B) with $n_p^{dest}$ publications starting from a university or country with $n_p^{orig}$ publications. The dark areas represent our data while the light ones a reshuffled model where the number of publications are assigned randomly to each node keeping unchanged the network topology. Indexes on the $x$ and $y$ axes represent aggregated data for all the nodes for which $2^j < n_p \le 2^{j-1}$. For a sake of clarity we represented only three significant cases for the origin publication ranges: $n_p^{orig} \in B_1, B_8, B_{16}$, for the universities and $n_p^{orig} \in B_1, B_{10}, B_{20}$, for the countries. Panels C: Average number of publications of the destination country as a function of the number of publications of the origin one. Panels D: Average number of publications of the destination university as a function of the number of publications of the origin one. The comparison with the randomised model in the upper plots shows that in real data a strong preference toward important university (or countries) is always present. For low publishing universities (or countries) we observe a larger spread of $n_p^{dest}$ and a lower level of connectivity with top publishing academies. In both cases this particular behaviour translates into the weak assortative tendency observed in the lower panel: universities with a small publication volume tend to be connected, on average, to less important, in terms of publications, nodes and viceversa. Panel E: Average number of publications of the destination country as a function of the number of publications of the origin one when both origin and destination are European countries (violet) and US (green) states, in case the movement is internal to the US. Panel F: Average number of publications of the destination university as a function of the number of publications of the origin one when both universities belong to an European country (violet) and US (green) in case the jump occurred between academies located in US.
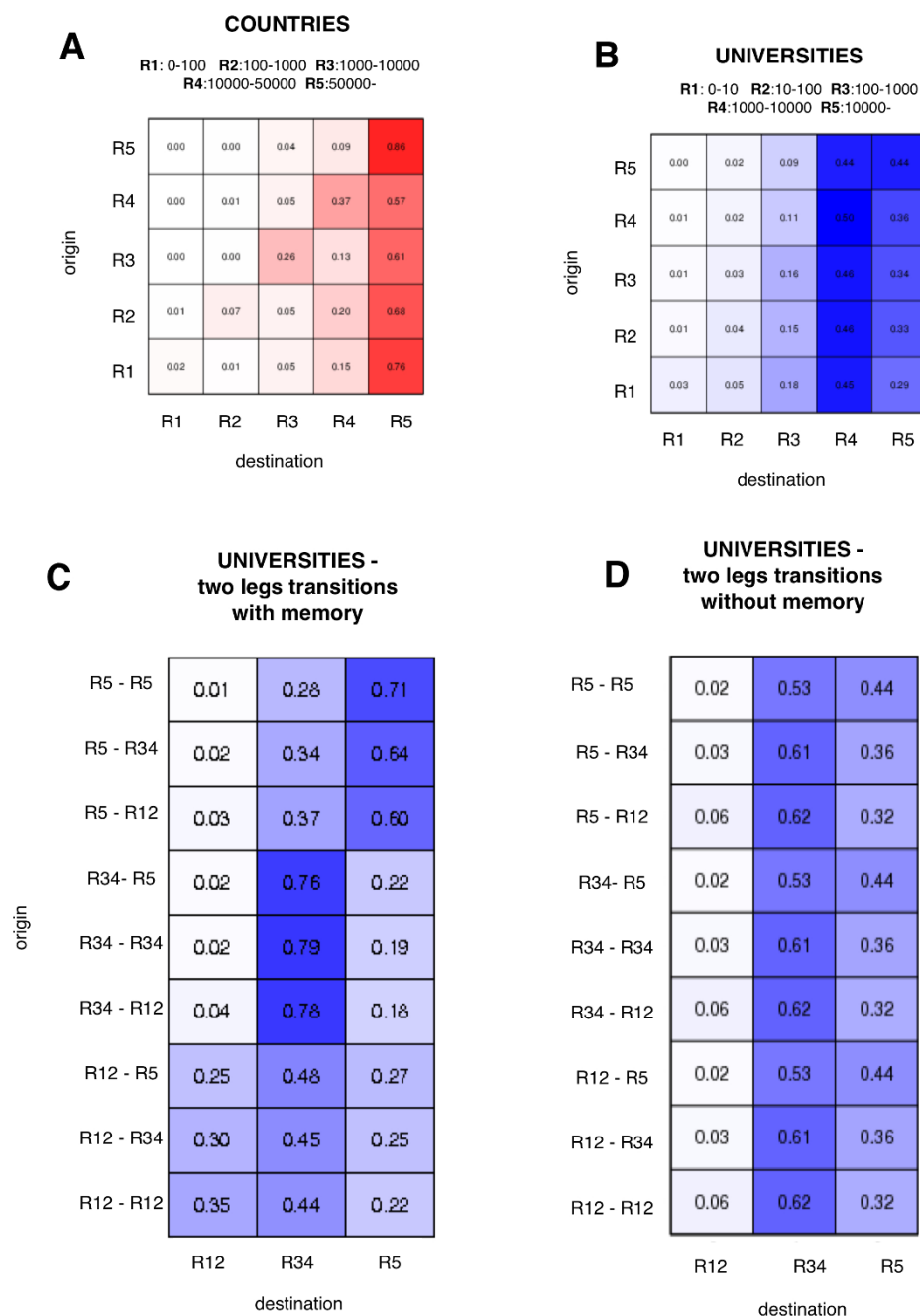
**Figure 7** | Panel A: probability transitions from each of the 5 ranked countries to each of the 5 destinations countries. Panel B: probability transitions from each of the 5 origins universities to each of the 5 destination universities. While we can observe that a large amount of movements are toward high ranked countries, in the case of universities most of the scholars trajectories tend to high but not to the highest universities, except for researchers already starting from a highest ranked university. Panel C: transition probabilities with memory. Plot D: transition probabilities without memory. Each value is given by the normalised product of the rank transitions on the two legs. All the possible 2 legs connections are considered, once we merge the first and the second ranked universities into a unique class R12, and the third and the fourth ones into a single one, R34. From such results we can conclude that the system exhibits a memory effect and thus the first affiliation will play an important role, moreover two consecutive affiliations in a low–medium university will prevent the scholar to reach a higher one, similarly whenever a researcher realises two stays in high ranked universities, her/his career will almost surely remains in the top group universities.

remain in a low rank one. On the contrary, starting from an high ranked university, the probability of ending in a low rank one is almost zero.

1. Cepii. Geographical distances database (last accessed on 16 june 2013 ) URL: http://www.cepii.fr.
2. APS physics. The American physical society site (last accessed on 8 september 2013 ) URL: http://www.aps.org.
3. Hirsch, J. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **102**, (2005).
4. Kaur, J., Radicchi, F. & Menczer, F. Universality of scholarly impact metrics. *J. Informetr.* **7**, 924–932 (2013).
5. Petersen, A. M., Riccaboni, M., Stanley, H. E. & Pammolli, F. Persistence and uncertainty in the academic career. *Proc. Natl. Acad. Sci. USA* **109**, (2012).
6. Wang, D., Song, C. & Barabasi, A.-L. Quantifying Long-Term Scientific Impact. *Science* **342**, 127–131 (2013).
7. de Solla Price, D. J. Networks of Scientific Papers. *Science* **149**, 510–515 (1965).

9

8. Newman, M. E. The structure of scientific collaboration networks. *Proc. Natl. Acad. Sci. USA* **98**, (2001).

9. Ramasco, J. J., Dorogovtsev, S. N. & Pastor-Satorras, R. Self-organization of collaboration networks. *Phis. Rev. E* **70**, 036106 (2004).

10. Pan, R. K., Kaski, K. & Fortunato, S. World citation and collaboration networks: uncovering the role of geography in science. *Sci. Rep.* **2**, (2012).

11. Cañibano, C., Otamendi, J. & Andùjar, I. Measuring and assessing researcher mobility from CV analysis: the case of the Ramòn y Cajal programme in Spain. *Res. Evaluat.* **17**, 17–31 (2008).

12. VV.AA. *The International Mobility of Talent: Types, Causes, and Development Impact* [Solimano, A. (ed.)] (Oxford University Press, 2008).

13. Ackers, L. Internationalisation, mobility and metrics: A new form of indirect discrimination? *Minerva* **46**, 411–435 (2008).

14. Gonzalez, M. C., Hidalgo, C. A. & Barabasi, A.-L. Understanding individual human mobility patterns. *Nature* **453**, 779–782 (2008).

15. Lambiotte, R. *et al.* Geographical dispersal of mobile communication networks. *Physica A* **387**, 5317–5325 (2008).

16. Krings, G., Calabrese, F., Ratti, C. & Blondel, V. D. Urban gravity: a model for inter-city telecommunication flows. *J. Stat. Mech. Theor. Exp.* **2009**, L07003 (2009).

17. Song, C., Qu, Z., Blumm, N. & Barabasi, A.-L. Limits of Predictability in Human Mobility. *Science* **327**, 1018–1021 (2010).

18. De Montis, A., Barthelemy, M., Chessa, A. & Vespignani, A. The structure of interurban traffic: a weighted network analysis. *Env. Plan. B* **34**, 905–924 (2007).

19. Lenormand, M., Huet, S., Gargiulo, F. & Deffuant, G. *A universal model of commuting networks.PLoS ONE* **7**, e45985 (2012).

20. Barrat, A., Barthelemy, M. & Vespignani, A. *Dynamical Processes On Complex Networks.* (Cambridge University Press, 2008).

21. Barrat, A., Barthelemy, M., Pastor-Satorras, R. & Vespignani, A. The architecture of complex weighted networks. *Proc. Natl. Acad. Sci. USA* **101**, (2004).

22. Barthélemy, M. Spatial networks. *Phys. Rep.* **499**, 1–101 (2011).

23. Rosvall, M., Esquivel, A. V., Lancichinetti, A., West, J. D. & Lambiotte, R. Networks with Memory. arXiv:1305.4807.

24. Radicchi, F., Fortunato, S. Markines, B. & Vespignani, A. Diffusion of scientific credits and the ranking of scientists. *Phis. Rev. E* **80**, 056103 (2009).

25. Radicchi, F. & Castellano, C. Rescaling citations of publications in physics. *Phis. Rev. E* **83**, 046116 (2011).

26. Duch *et al.* The Possible Role of Resource Requirements and Academic Career-Choice Risk on Gender Differences in Publication Rate and Impact. *PLoS ONE* **7**, e51332 (2012).

## Author contributions

F.G. and T.C. performed simulations and data analysis and wrote the paper.

## Additional information