
This is an electronic reprint of the original article.
This reprint may differ from the original in pagination and typographic detail.

Pan, Raj Kumar; Sinha, Sitabhra; Kaski, Kimmo; Saramaki, Jari

The evolution of interdisciplinarity in physics research

Published in:
Scientific Reports

DOI:
[10.1038/srep00551](https://doi.org/10.1038/srep00551)

Published: 01/01/2012

Document Version
Publisher's PDF, also known as Version of record

Published under the following license:
CC BY-NC-SA

Please cite the original version:
Pan, R. K., Sinha, S., Kaski, K., & Saramaki, J. (2012). The evolution of interdisciplinarity in physics research. *Scientific Reports*, 2, 1-8. Article 551. <https://doi.org/10.1038/srep00551>

This material is protected by copyright and other intellectual property rights, and duplication or sale of all or part of any of the repository collections is not permitted, except that material may be duplicated by you for your research use or educational purposes in electronic or print form. You must obtain permission for any other use. Electronic or print copies may not be offered, whether for sale or otherwise to anyone who is not an authorised user.



The evolution of interdisciplinarity in physics research

Raj Kumar Pan¹, Sitabhra Sinha², Kimmo Kaski¹ & Jari Saramäki¹

¹Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland, ²The Institute of Mathematical Sciences, C.I.T. Campus, Taramani, Chennai - 600 113 India.

SUBJECT AREAS:
STATISTICAL PHYSICS,
THERMODYNAMICS AND
NONLINEAR DYNAMICS
GENERAL PHYSICS
STATISTICS
MODELLING AND THEORY

Received
24 May 2012

Accepted
19 July 2012

Published
1 August 2012

Correspondence and
requests for materials
should be addressed to
R.K.P. (raj कुमार.pan@
aalto.fi)

Science, being a social enterprise, is subject to fragmentation into groups that focus on specialized areas or topics. Often new advances occur through cross-fertilization of ideas between sub-fields that otherwise have little overlap as they study dissimilar phenomena using different techniques. Thus to explore the nature and dynamics of scientific progress one needs to consider the organization and interactions between different subject areas. Here, we study the relationships between the sub-fields of Physics using the Physics and Astronomy Classification Scheme (PACS) codes employed for self-categorization of articles published over the past 25 years (1985–2009). We observe a clear trend towards increasing interactions between the different sub-fields. The network of sub-fields also exhibits core-periphery organization, the nucleus being dominated by Condensed Matter and General Physics. However, over time Interdisciplinary Physics is steadily increasing its share in the network core, reflecting a shift in the overall trend of Physics research.

Scientific progress has been seen both as a succession of incremental refinements as well as a succession of epochs with relatively slow or little change that are punctuated by periods of revolutionary transitions. In Popper's view¹, science proceeds by gradually falsifying competing candidate theories, whereas Kuhn² argues that during episodes of “normal science”, scientists gradually improve their theories within the current framework until enough unexplainable anomalies emerge to call for a major paradigm shift. Such shifts have occurred on many scales, from scientific revolutions with global reverberations to smaller breakthroughs within specific fields or sub-fields of science. However, this view ignores the possibility of entirely new avenues of research emerging from new connections that are forged between apparently disjoint areas of science. Thus, new paradigms may be born not only because of evidence that contradicts existing theories, but also because entirely new questions and theoretical frameworks appear. For example, consider the rise of systems biology, driven by technological advances in data acquisition and their analysis through computer algorithms, or the emergence of network science that merges aspects from physics, computer science, and social sciences.

In this paper, we focus on the dynamics and emergence of connections between the various subfields of physics, and perform a longitudinal analysis of the evolution of physics from 1985 till 2009. Our results are based on a study of the papers appearing in the Physical Review series of journals (Physical Reviews A, B, C, D, E, Physical Review Letters and Review of Modern Physics) published by the American Physical Society during this period, with their Physics and Astronomy Classification Scheme (PACS) numbers indicating the subfields of physics to which they belong. If a paper is listed under two different PACS codes, the two corresponding sub-fields are considered to be connected by the paper. In this manner we construct a set of annual snapshots of the networks of sub-fields in physics that are connected through all papers that have been published in each year, and study the evolution of these networks at multiple structural scales. In this way, we can focus on the big picture of the evolution of physics in terms of changes in the nature of connections between its subfields, instead of the microscopic level that is considered by the widely studied collaboration or citation networks^{3–6}.

We show that the network of the subfields of physics is becoming increasingly connected over time, both in terms of link density and the numbers of papers joining different subfields. Despite gradual changes in the network density, composition, and degrees of individual nodes, all key statistical distributions display scaling, indicating stationarity in the underlying micro-dynamics⁷. It is seen that a substantial and increasing fraction of new links connects nodes that belong to dissimilar branches of the PACS hierarchy, reflecting a trend where interdisciplinarity between the subfields of physics clearly increases. By applying the *k*-shell decomposition technique, we show that the core of physics has been dominated by Condensed Matter and General Physics for the entire period under study, with Interdisciplinary Physics steadily increasing its importance in the core. It is seen that a substantial and increasing fraction of new links connects nodes that belong to dissimilar branches of the PACS hierarchy, reflecting a trend where interdisciplinarity between the subfields of physics clearly increases. By



applying the k -shell decomposition technique, we show that the core of physics has been dominated by Condensed Matter and General Physics for the entire period under study, with Interdisciplinary Physics steadily increasing its importance in the core.

Results

We have analyzed all published articles in Physical Review (PR) journals⁸ from 1985 till the end of 2009 which are classified by their authors as belonging to certain specific sub-fields using the corresponding PACS codes. The PACS is an internationally adopted, hierarchical subject classification system of the American Institute of Physics (AIP) for categorizing publications in physics and astronomy⁹. It is primarily divided into 10 top-level categories that represent broad research areas. Each of these categories are then divided into smaller domains representing more specific fields of physics, which may be further split into even more specific sub-fields. Thus, each of these PACS codes represent a specific sub-field of physics. (for a detailed description of the data, see Methods). For constructing the networks of the different sub-fields, we consider the PACS codes as nodes, a pair of which are linked if an article is classified by both these codes. In these networks, the degree k of a node corresponds to its number of links, *i.e.* number of other PACS codes it is connected to, and its strength s to the total number of articles published with the PACS code. The numbers of papers sharing two PACS codes are accounted for with the weight w of their link. In order to study the time evolution of this system, we create yearly aggregated networks by considering all the articles published in a given year (see Methods).

Network-level evolution of the system. We begin by considering the evolution of the overall system properties between 1985 and 2009. For these 25 years, the total number of yearly publications N_{papers} in all PR journals has grown linearly [Fig. 1(a)], while the number of PACS codes N_{PACS} shows a linear increase between 1990 and 2002,

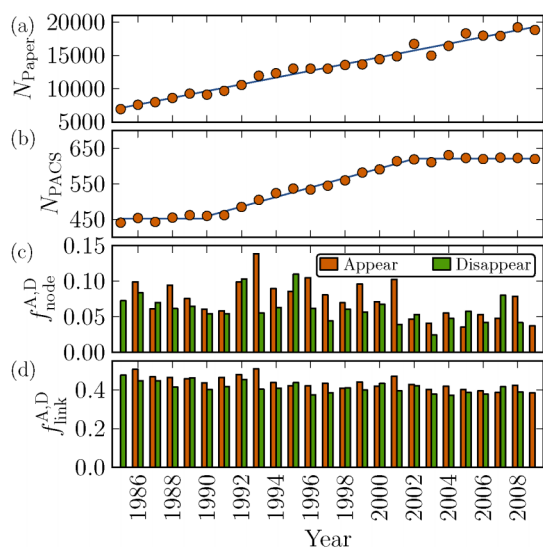


Figure 1 | The time evolution of various properties of the PACS network: (a) the number of published papers, (b) the number of PACS codes, (c) the fraction of new and disappeared nodes, (d) the fraction of new and disappeared links, (e) the average degree, $\langle k \rangle$, and (f) the average link weight, $\langle w \rangle$. The solid lines in (a), (e) and (f) denote a linear growth of $\langle \Delta N_{\text{papers}} \rangle \approx 508$ papers per year, a yearly increase of $\langle \Delta k \rangle \approx 0.44$ of the average degree $\langle k \rangle$, and a yearly increase of $\langle \Delta w \rangle \approx 0.02$ of the average link weight $\langle w \rangle$, respectively. The solid line in (b) shows two roughly constant regimes, interspersed by a period of average linear increase of $\Delta N_{\text{PACS}} = 13.5$ PACS codes per year between 1990–2002. Note that k and w are heterogeneously distributed; see Fig. 2.

remaining roughly constant before and after this period. Note that this does not imply that the same codes have been in use in all the years prior to 1990 or those after 2002, but rather that the number of new PACS codes that were introduced each year were approximately balanced by the number of codes that were discontinued that year. The fraction of new and removed PACS codes each year is seen to fluctuate between 5% and 15% in Fig. 1(c). The yearly fractions of new and disappearing links between PACS codes are higher, fluctuating around $\sim 40\%$ [Fig. 1(d)]. When looking at network averages of the degree $\langle k \rangle$ and link weight $\langle w \rangle$ [Fig. 1(e),(f)], it is seen that not only does the number of published papers grow, but the network also becomes more connected, as both $\langle k \rangle$ and $\langle w \rangle$ grow approximately linearly. As a consequence, the average path length of the network decreases linearly (see Supplementary Information). Thus, in general, the connectivity between different subfields of physics is increasing with time.

The scaled cumulative distributions of the key quantities (degree k , strength s , and link weight w) are shown in Fig. 2 for four different years. All distributions are broad and indicate heterogeneity – compared to the averages, some subfields of physics are much more connected to the rest, the links between some fields are stronger, and many more papers are published in some fields. Furthermore, the overlap of the rescaled distributions indicates that although the averages of the distributions are growing over time, the functional form of the distributions remains similar^{7,10}. This is corroborated by comparing the Kolmogorov-Smirnov (KS) statistic of the degree distribution of the yearly networks with each other and finding that the KS distance stay at a low constant value¹¹. A similar comparison of the KS statistics of the strength distribution of the yearly networks shows similar behavior, although there is a slight deviation from this general pattern for the year 1985 (see Supplementary Information for further details). Hence, although the composition of the system changes over time in terms of nodes and links appearing and disappearing (Fig. 1), the functional shape of the key distributions remain similar across the years, indicating stationarity at the level of macro dynamics.

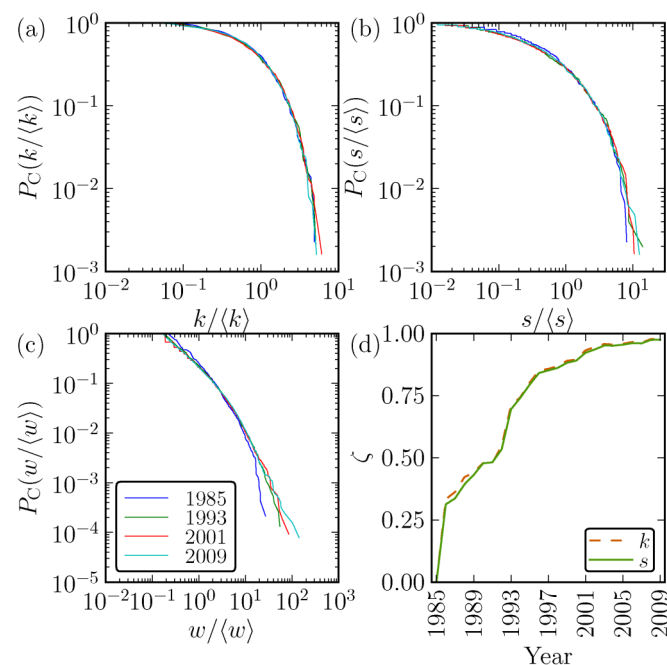


Figure 2 | Stationarity of the macro-level statistical distributions and variation at the micro level with time. The cumulative distributions of (a) degree k , (b) strength s and (c) link weight w of the PACS network, for four different years. The curves have been scaled by their averages for the given year. (d) The dissimilarity coefficient ζ for the degree and strength ranks of the nodes, between the year 1985 and subsequent years.



In contrast to the relative invariance of the distributions, we observe that over a long timescale the degrees and strengths of some nodes in the network increase or decrease in rank over time. Fig. 2(d) displays the dissimilarity coefficient ζ of the degree ranks¹² (see Methods) with respect to the year 1985 as a function of time; $\zeta \in [0, 1]$ such that low values indicate invariant node ranks. It is seen that ζ increases monotonically with time, approaching $\zeta \approx 1$ towards the end. Thus, the degree ranks of the PACS codes change gradually over time and become uncorrelated towards the end of the period under study, indicating the presence of longer-term trends. Using the node strength to calculate ζ or calculating ζ between all pairs of years yields similar results (see Supplementary Information). We also compare the structural properties of the empirical PACS network with a randomized ensemble, in which PACS codes are reshuffled among papers. This is to see whether the observed properties of the network are expected to appear purely by chance as a consequence of the constraints inherent in the system. We found that in the randomized version there are many more links in the network compared to the empirical network leading to an increase in the clustering coefficient, decrease in the average link weight, and decrease in the average path length (see Supplementary Information).

Micro-level dynamics. Next we take a detailed look at the micro-dynamics of new and disappearing links and nodes. We take advantage of the hierarchical nature of the PACS scheme (see Methods), and consider the hierarchical similarity h of two PACS nodes. Nodes are considered dissimilar ($h = 0$), if they belong to different main branches of the PACS hierarchy and thus represent very different subfields of physics. Nodes can also represent related subfields of physics and be similar with respect to the first level of hierarchy ($h = 1$, i.e., they share their first PACS digit), or similar with respect to the second level ($h = 2$, i.e., they are even more similar since they share the first two PACS digits). First, we focus on the link density ρ of the network, defined for each similarity class as the number of links between nodes of the class normalized by the number of pairs of nodes in the class. The evolution of the link density between dissimilar nodes ($h = 0$) and nodes belonging to the same second hierarchical level ($h = 2$) is displayed in Figs. 3(a) and (b). For both cases, the density increases with time. As one would expect, the link density for $h = 2$ nodes is far higher than that between dissimilar nodes. However, the relative increase of the density between the $h = 0$ nodes is much higher, indicating an increasing trend where new connections emerge between the main branches of physics. If the new links of each year are split into fractions according to whether they connect similar or dissimilar sub-fields [Fig. 3(c–d)], it is seen that a substantial and increasing fraction of new links connects nodes that belong to dissimilar branches of the PACS hierarchy ($h = 0$), while the fraction of new links joining similar PACS codes ($h = 2$) decreases with time. Thus, there is an increase in interdisciplinarity between the subfields of physics, as dissimilar branches of the PACS hierarchy are becoming increasingly connected. This result holds even with a randomized null model that takes into account the different numbers of $h = 0$ and $h = 2$ nodes (see Supplementary Information). Furthermore, this hierarchical connectivity and the increase in the interdisciplinarity of the empirical network is lost in a randomized network constructed by randomly shuffling the PACS codes across different papers (see Supplementary Information).

Let us next address the role of network topology in the micro-dynamics. In particular, we want to see whether new links reflect the clustered structure of the network, increasing the density of dense neighborhoods as exemplified by the visualization of Fig. 4(a). Additionally, since the PACS numbers themselves evolve and new codes appear, local clusters may also become increasingly connected if new nodes joining nearby nodes appear, as in Fig. 4(b). The disappearance and appearance of nodes may

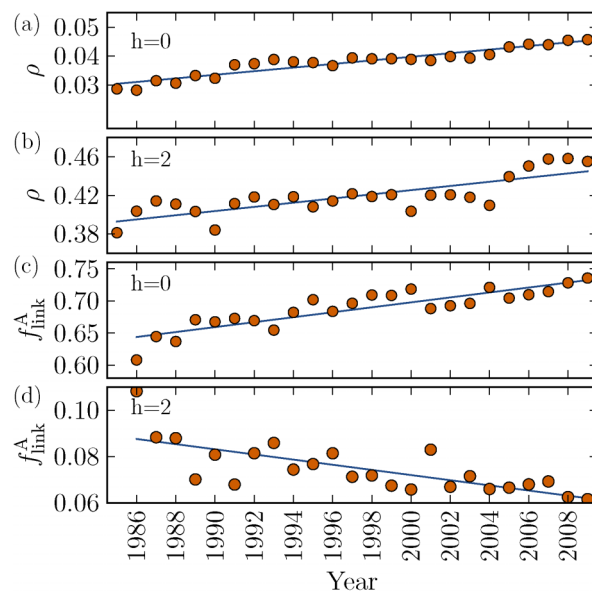


Figure 3 | The time evolution of network density and newly appearing links. Evolution of the link density ρ within the sets of nodes that are hierarchically (a) dissimilar, and (b) similar up to second level. Time dependence of the fraction of new links f_{links}^A that connect nodes that are hierarchically (c) dissimilar and (d) similar up to the second PACS level. The solid curves indicate linear increase in (a), (b) and (c) with slope 6.2×10^{-4} , 2.2×10^{-3} and 3.9×10^{-3} , respectively and linear decrease in (d) with slope 1.1×10^{-3} .

also reflect structural changes in the PACS system, such as code replacement [Fig. 4(c)].

First, we look for evidence for the mechanisms of Fig. 4(a) and (b), where new links are not randomly created, but follow a process where dense clusters of interlinked PACS codes become even denser. For this, we determine the geodesic distance d (the number of links on the shortest path) and the number of common neighbors n_{CN} for all pairs of nodes for each year, and count the number of pairs that are joined through a new link or through a new intermediate node in the following year. This allows us to calculate the probabilities of link and connecting node appearance ($P_{\text{links}}^A, P_{\text{node}}^A$) aggregated over the data interval. Their dependence on the geodesic distance and number of common neighbors is shown in Fig. 4 (d)–(g), where we have further divided all node pairs into PACS similarity classes ($h = 0, 1, 2$ as above). It is evident that the closer the nodes are and the more common neighbors they have, the higher the likelihood of the appearance of a new direct link or a new joint neighbor connecting the nodes. The mechanisms of Figs. 4(a) and (b) are thus common in the network, and new connections between the sub-domains of physics do not emerge in a random, uncorrelated fashion; rather, connectivity increases within clusters. Furthermore, the more similar a pair of nodes is with respect to the PACS hierarchy, the higher the likelihood of new connections between them. Similar features have also been seen in other networks, e.g., in social networks new links are more likely to appear between nodes that are close, that is, nodes that have common friends or share similar interests^{12–14}.

In order to study code replacement dynamics of Fig. 4(c), where discontinued codes are replaced by new codes that have a similar connectivity pattern, we define a weighted version of the neighborhood overlap O_{ij}^w between a pair of nodes. This overlap is used to determine the similarity in the neighborhood of two nodes so that $O_{ij}^w = 0$ if nodes i and j have no common neighbors, and $O_{ij}^w = 1$ if they have same set of common neighbors (see Methods). We study all PACS codes that have been discontinued, and first find their peak years t^* with the highest number of papers. For each PACS code i , we

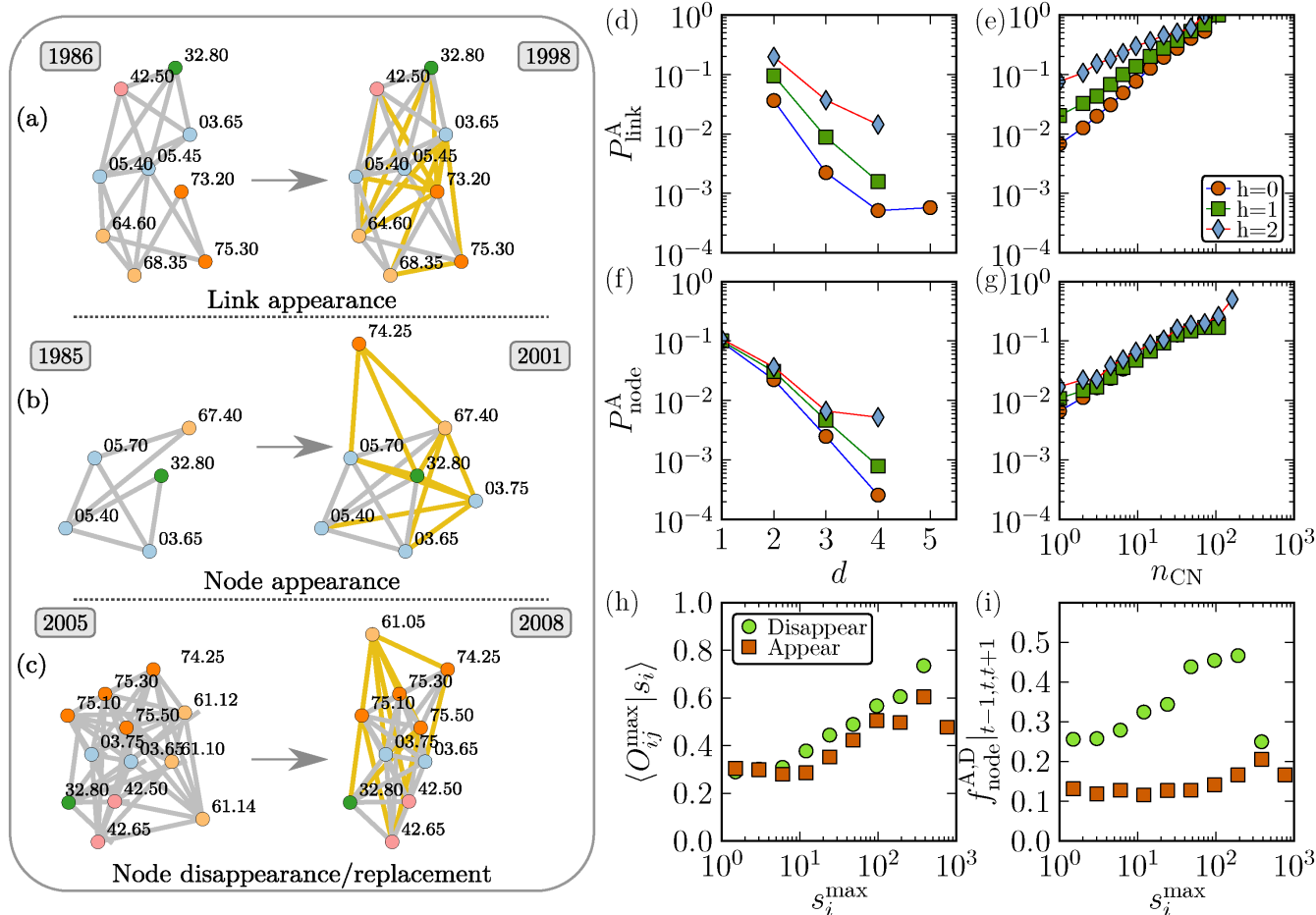


Figure 4 | Micro-dynamics in the PACS network. Examples of (a) the appearance of new links that increase the density of a local neighborhood: a large number of links appear between some sub-fields of condensed matter and general physics, (b) the appearance of new nodes (03.75, 74.25) and links increasing the density and (c) changes in the network structure, where a new node (61.05) replaces several existing nodes (61.10, 61.12, 61.14). Probability of link appearance as a function of the (d) distance d and (e) number of common neighbors n_{CN} between the nodes. Probability of appearance of new nodes as a function of the (f) d and (g) n_{CN} between the nodes. The links categorized according to the hierarchical similarity h of the nodes they are connecting. (h) Similarity of discontinued nodes (circles) and newly introduced nodes (squares) with their maximally similar counterpart nodes, as measured by the overlap O_{ij}^{\max} . The overlap is averaged over focal node strength. (i) The fraction of maximally similar nodes that appear around the disappearance of the focal node (circles) (from one year before the disappearance to one year after) and the fraction of maximally similar nodes that have disappeared around the time of appearance of the focal node (squares), again as a function of focal node strength.

determine the network neighborhood Λ_{i,t^*} corresponding to the peak year. We then calculate the overlap of this neighborhood with the neighborhoods of all nodes in the network at year $t_i + 1$, where t_i is the year when i becomes discontinued. We then choose the node j whose link pattern has the closest match with i at its peak, as indicated by the maximum overlap with Λ_{i,t^*} . The average of this maximum overlap O_{ij}^{\max} is displayed as a function of the strength of the disappearing nodes s_i in Fig. 4(h). The overlap increases with the strength of the discontinued node. Thus for high-strength nodes, nodes of similar neighborhoods are present immediately after their disappearance. These similar nodes are also usually introduced around the time of discontinuation (see Fig. 4(i)). Hence high-strength PACS codes frequently get replaced rather than disappear altogether; this can be taken indicative of gradual, continuous changes in the subfields of physics. This might be due to the changing perceptions about sub-fields as a result of gradually improving understanding of their place in the general scheme of physics. These newly appearing codes have connectivity similar to the disappearing PACS and also have many new connections to other different sub-fields.

When a similar analysis is performed focusing on PACS codes that are newly introduced, it is seen that nevertheless, the majority of new

codes correspond to emerging new subfields and do not appear to replace existing codes (see Supplementary Information).

Mesoscopic structure. *The maximum spanning tree.* We now shift our focus from micro-dynamics towards the mesoscopic level and begin by illustrating the structure of the PACS network with the help of its maximum spanning tree (MST). The MST is a tree connecting all nodes of the network while maximizing the sum of link weights; such trees can be used to explore structural features in the data (see, e.g., Ref. 15). Figure 5 displays the MST for the PACS network of the year 2009 (874 nodes). Some structural features are apparent: first, as expected, PACS codes belonging to the same broad categories are frequently connected in the MST; however, there is mixing as well, especially in the central parts of the tree. Second, the MST reflects the underlying cluster structure of the network. There appears to be a branch that is well separated from the rest, containing fields related to high-energy physics: Physics of Elementary Particles and Fields, Nuclear Physics, and Geophysics, Astronomy and Astrophysics. The rest of physics displays more mixing in the MST, the hub nodes being frequently related to General Physics, Optics, and Condensed Matter.

k^{-shell} analysis. Although the minimum spanning tree visualization of the network provides some overview on the structural



organization of the relations between the different subfields of physics, it neither indicates the significance of the nodes forming the core of the network nor gives us any information regarding the temporal evolution of the structure. For a better and more detailed understanding, we perform k -core analysis^{16–19} of the evolving PACS network by decomposing the network for each year into its k^s -shells (see Methods), such that a high k^s -shell index of a node reflects a central position in the core of the network.

First, we want to establish that the k^s -shell indices of the PACS codes are relatively stable over time and are thus suitable for analysis. To do this we determine the correlation coefficients between the k^s -shell indices of all the PACS codes and between different years. In Fig. 6 (a) the correlation coefficient between different pairs of years are represented in terms of a matrix with the color of each cell representing the corresponding correlation value. The coefficient has a high value for neighboring years, so that changes in the shell indices of nodes appear gradual over time rather than randomly. Thus, the nodes having high or low k^s -shell index for year t are more likely to retain their index for the subsequent year $t + 1$. Furthermore, the correlation matrix shows a block diagonal structure, indicating higher correlations for three periods, 1985–1992, 1993–2000 and 2001–2009. For analysis of k^s -shell regions (see below), we pick one network corresponding to each of these periods. The k^s -shell indices of PACS codes are also related to their stability. We define a node as stable if it has been in use each year after its introduction. Fig. 6 (b) shows the fraction of stable nodes calculated over the entire period 1985–2009 as a function of the k^s -shell index; it

is evident that the higher the order of the k^s -shell (and thus, the closer it is to the nucleus of the network), the larger is the fraction of stable nodes. Note that, as the k^s -shell index of a node is related to its degree and strength, nodes that have high degree or strength are also less likely to get deleted and are more stable.

For studying the time evolution of the k^s -shells, we use the alluvial diagram method²⁰. We divide the PACS codes into four categories based on their k^s -shell indices by dividing the range of k^s values into four groups of approximately equal sizes. Thus Region I contains codes that are in the core of the network ($k^s \in [\frac{3}{4}k_{\max}^s, k_{\max}^s]$), and

Regions II, III, and IV contain nodes with increasingly lower k^s -shell indices. The colored blocks of the alluvial diagram in Figure 7 show the different regions for three different years, with the size of each block representing the number of PACS codes in the respective region. The sizes are increasing with time, indicating an increase in the number of PACS codes. Furthermore, the maximum shell index k_{\max}^s has increased with time, as indicated by the color of the k^s -shell indices for different years.

The shaded areas joining the k^s -shell regions represent flows of PACS codes between the regions, such that the width of the flow corresponds to the fraction of nodes. The total width of incoming flow is less than the width of the corresponding region, because the rest is made up by new PACS codes entering the network. Likewise, the gap between the width of the block and total outgoing flow corresponds to discontinued PACS codes. Here, it is seen that the core of the network, Region I, is remarkably stable compared to the

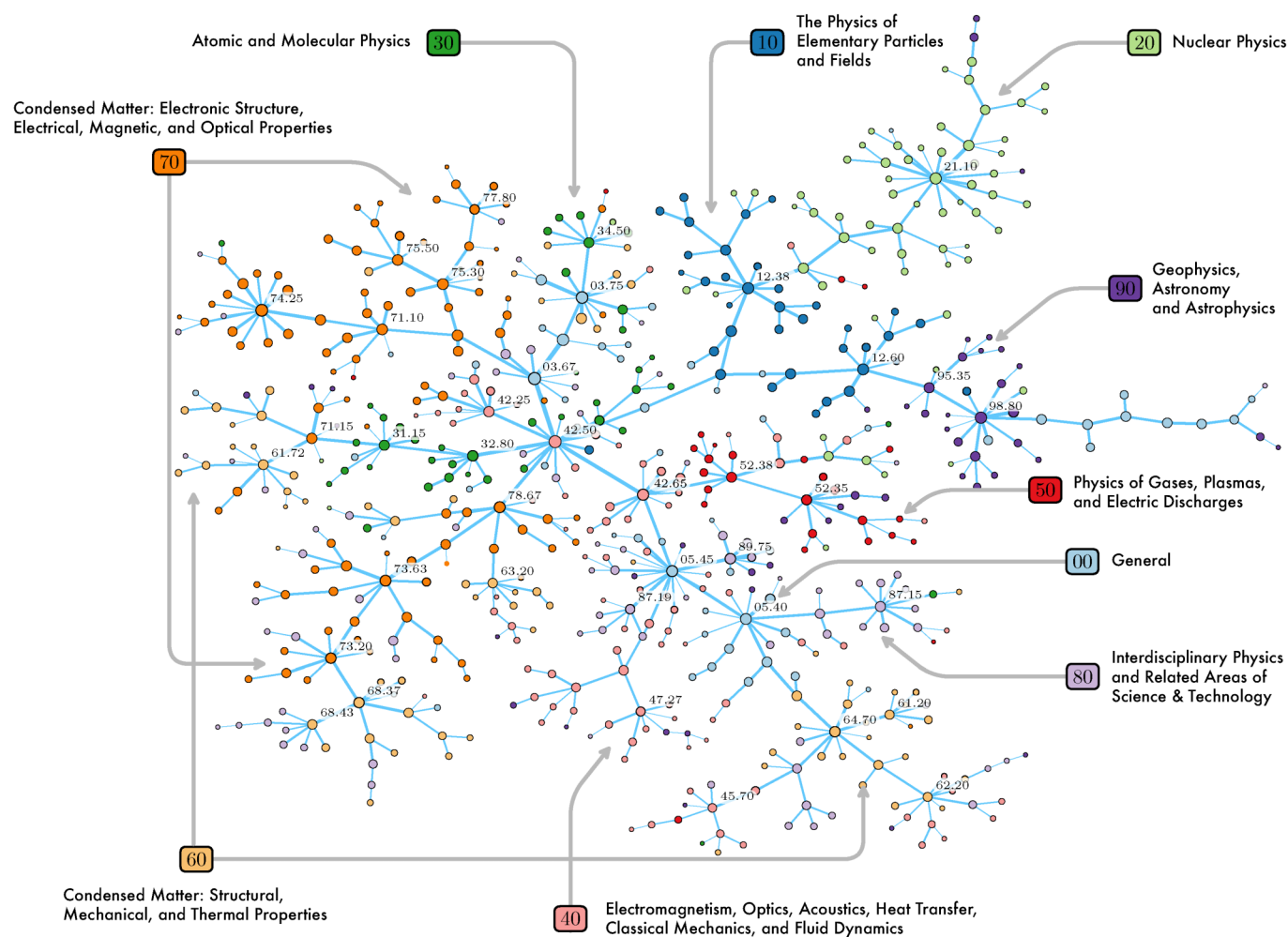


Figure 5 | The maximum spanning tree of the PACS network of 2009.

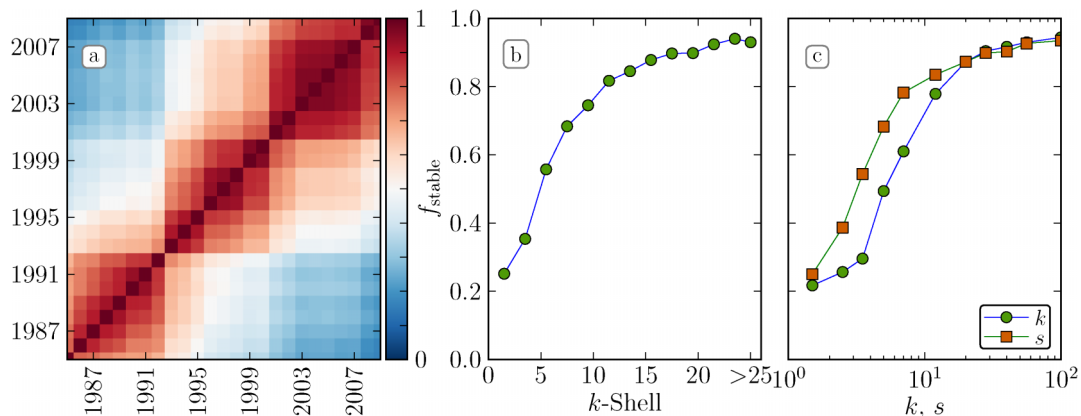


Figure 6 | (a) Matrix showing the correlation coefficient between the k^i -shell indices of the PACS codes for different years. The fraction of the stable PACS codes that have remained present since their introduction as a function of its (b) k^i -shell index. (c) degree and number of appearances.

peripheral Region IV that displays a high turnover of codes. Nodes that are in the core of the network are highly likely to remain so, whereas peripheral nodes frequently either disappear or migrate towards the core. Furthermore, a high fraction of new nodes first appear in the peripheral region.

Next, we consider how the different branches of physics are positioned with respect to the core-periphery organization of the PACS network and how their position has changed over time. Figure 8 displays multi-level pie charts for three different years, where each level of the chart represents one of the k^i -shell regions as above. The innermost layer represents Region I, followed by Region II, Region III, and finally the outermost layer represents the peripheral Region IV. For each layer, we show the fraction of level-3 PACS codes belonging to the different branches of physics as indicated by their first hierarchical PACS level.

The pie chart for the year 1987 shows that the core region I consists mostly of General Physics and Condensed Matter (PACS categories

00, 60 and 70), with a small contribution from categories 30 (Atomic and Molecular Physics), 40 (Electromagnetism etc), and 80 (Interdisciplinary Physics). In all other regions, all branches of physics are present. For the network structure of 1997, we see that the contributions of PACS categories 30, 40, and 80 have increased in the core region. Looking at the pie chart for the year 2007, we see that Interdisciplinary Physics (80) has taken over an even larger fraction of the core. The three main groups in the core are the two Condensed Matter categories (60, 70) and Interdisciplinary Physics (80). At the same time, it is seen that Nuclear Physics (20) has been moving towards the periphery, mainly contributing to Region III; this is in line with its position in the MST of Fig. 5. Thus, between 1987 and 2009, we see that Condensed Matter and General Physics have retained their position in the very core of physics, while Interdisciplinary Physics has been steadily moving towards the core, and Nuclear Physics has migrated towards the periphery. Furthermore, Physics of Elementary Particles and Fields (10) and

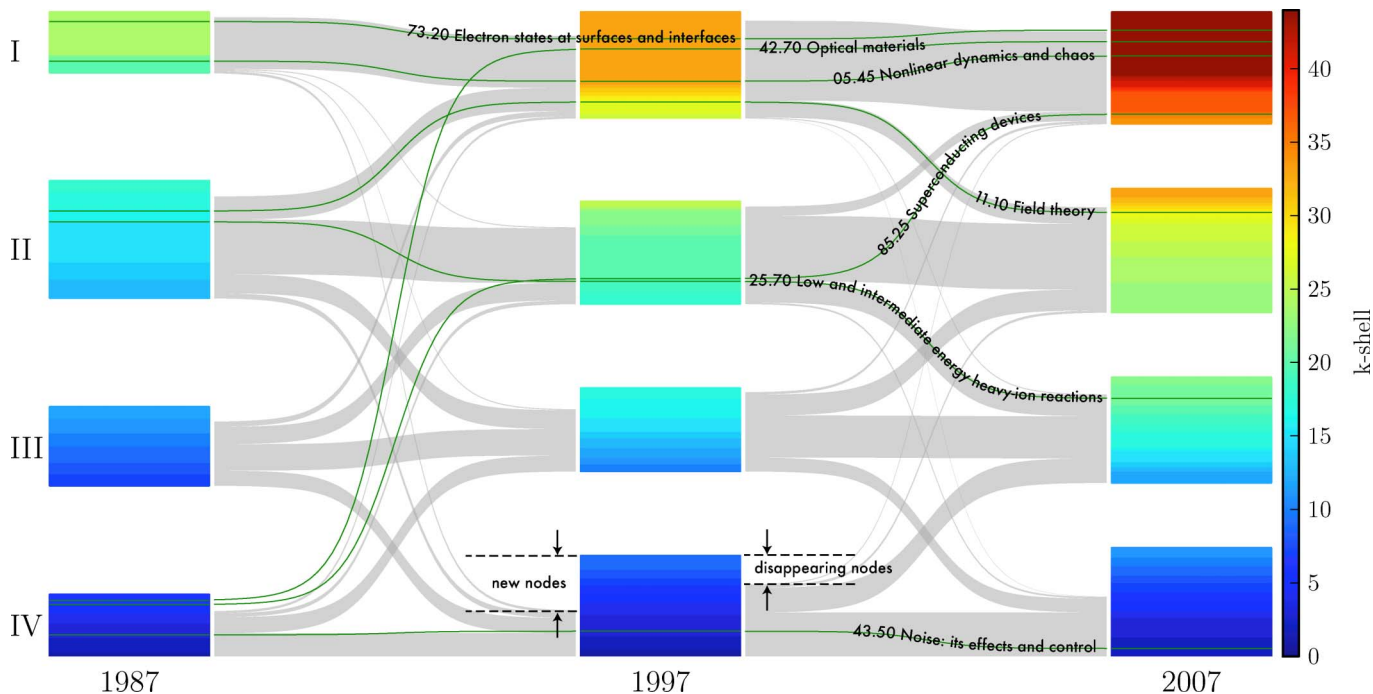


Figure 7 | Evolution of the k^i -shell indices of the PACS codes and the flows between k^i -shell regions between the years 1987, 1997, and 2007. The PACS codes for each year are divided into four different categories according to their k^i -shell index (indicated by the color). The size of the block indicates the number of codes in that category, and the widths of the shaded areas correspond to the fraction of migrating codes. The green lines show k^i -shell trajectories for some specific PACS codes as examples.

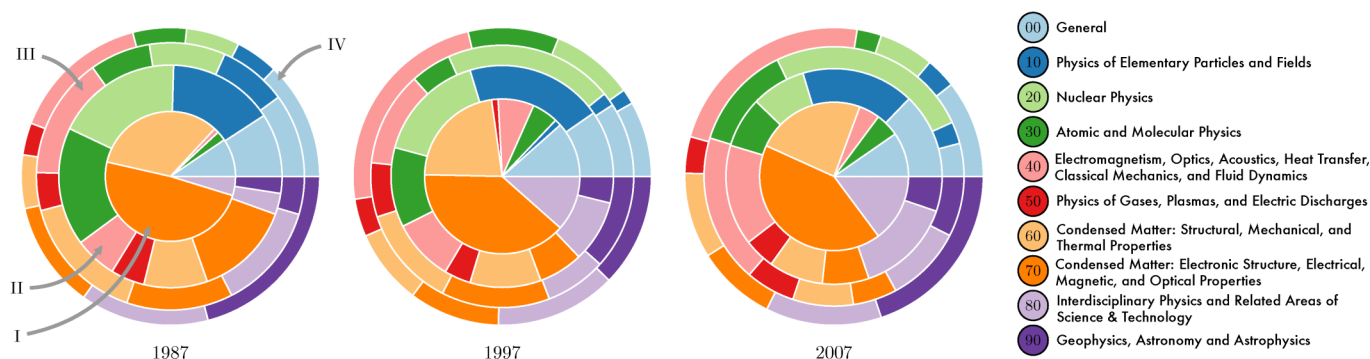


Figure 8 | Multi-level pie chart for year 1987,1997 and 2007 showing the composition of each of the PACS k -shell regions (I–IV), such that the colors represent the first level of the PACS hierarchy.

Astrophysics (90) have retained their relative core position during this period. Note that if the above pie charts are calculated on the basis of the total number of papers for each PACS code (see Supplementary Information), no clear evolution can be observed, as the codes are more homogeneously distributed in the regions. This indicates that within each hierarchical level-1 category, there are level 3 PACS codes with highly varying volumes of publication activity and this volume does not directly correspond to the position of the code in the network.

Discussion

We have studied the evolution of physics research in terms of inter-connections between its subfields from 1985 to 2009. We have shown that for yearly networks constructed from PACS codes, although there are apparent dynamical changes in the network, the key statistical distributions display remarkable stationarity. The average number of links per code and average link weight show a steady increase, indicating increased connectivity between different subfields of physics. In particular, the rate of link formation between subfields that are distant in the PACS hierarchy has increased, pointing out a clear trend of increased interdisciplinarity within physics where its different branches are becoming increasingly interlinked. This evolution does not appear random or uncorrelated; rather, within the branches there are subfields that are joined together in clusters, and there is a tendency where subfields in such clusters get connected through new links or new intermediate subfields with a high rate. The “mesoscopic” or intermediate-scale analysis of the network suggests an evolution towards increasing interdisciplinarity in physics, and a detailed study of the properties of such growing clusters would likely provide important insights into the evolution of physics.

At the mesoscopic level of the network, k -shell decomposition analysis reveals some largescale trends within physics discipline: the nodes in the core of the network display the highest probability of survival, whereas the peripheral region displays the largest turnover associated with the discontinuation of older PACS codes and the appearance of new ones, as well as, their migration towards the core. The nodes that are in the core have a large number of connections to a large number of other nodes, and thus a high k -shell index can be taken as indicative of the importance of a PACS code compared to the “rest” of physics. With this interpretation it is natural that such high- k -shell subfields of high importance are also subfields of high stability. In our data, the core of the network has been dominated by those PACS codes that belong to the main branches of Condensed Matter and General Physics for the entire period under study. However, we also note that there is an important trend of the PACS codes belonging to Interdisciplinary Physics to steadily migrate towards the core, so that at present these already occupy a significant fraction of the core.

In conclusion, there has been an increase in the interdisciplinarity within physics, as indicated by the evolution of interconnections

between different branches of physics. In addition there is an increase in the importance of Interdisciplinary Physics that also has connections to fields outside physics, as indicated by its share of the core in the PACS network. Although it may be easy to identify candidate drivers for this evolution, like the availability of vast amounts of digital data in several areas (e.g., financial markets, social systems) and an increasing number of problems requiring specialists from several fields within and outside physics (e.g., problems related to energy, climate, and biophysics), assessing their importance is beyond the scope of this study. It would be especially interesting to see how the availability of research grants in different sub-fields of physics correlate with our observations, and whether the evolution of physics follows the amount of funding available for its sub-areas or vice versa. This would require data about science funding collated from many sources. In addition, the PACS codes represent only one possible way to define the subfields of physics. Furthermore, there may be delays between developments in physics and respective changes in the PACS hierarchy. Nevertheless, we feel that it would be very interesting to compare our results with a study of the network of inter-relations between physics sub-fields constructed by using some other data than the PACS codes and recent methods such as community structure analysis of citation or co-authorship networks used to define the subfields.

Methods

Data description. A PACS code contains three elements: a pair of two-digit numbers separated by “.” and followed by two characters that may be lower- or upper-case letters or “+” or “-” signs. The first digit of the first two-digit number denotes the main category out of the 10 broad categories specified at the first level and the second digit gives the more specific field within that category. The second two-digit number specifies a narrower category within the field given by the first two digits. The last two characters may specify even more detailed categories up to the fifth level of hierarchy. As an example, in the PACS code 05.45.-a, the first digit “0” indicates “General”, adding the second digit “05”, denotes “Statistical physics, thermodynamics, and nonlinear dynamical systems” and 05.45.-a indicates “Nonlinear dynamics and chaos”; the “-” sign denotes the presence of one more level of hierarchy. Our source data comes in the form of the PACS codes of all published articles in Physical Review (PR) journals⁹ of the American Physical Society from 1985 till the end of 2009. In this study we use the PACS codes up to the third level of hierarchy, *i.e.*, only the first four digits of the PACS codes. This is a good choice for longitudinal analysis: at the third level of hierarchy, the PACS codes represent the subfields of physics well and all PACS codes that have been listed in the papers extend at least to this level. Furthermore, there are more fluctuations in the deeper levels – the PACS codes change over time, as the classification scheme is regularly revised by AIP.

Network construction. For constructing the networks, we consider the individual PACS codes as nodes, such that links between them indicate that they have appeared in the same article. In order to follow the time evolution of this system, we create yearly aggregated networks by considering all articles published in a given year. We then extract the largest connected components (LCC) for all the yearly aggregated PACS networks; all network properties in this paper have been calculated for LCCs. For all years, the LCC’s correspond to almost the whole network (> 99.5%).

The weight of the link between the PACS code nodes i and j is defined as $w_{ij} = \frac{1}{\sum_p n_p - 1}$, where the sum runs over the set of papers in which the PACS codes i and j appear together, and n_p is the number of PACS codes used in paper p . This ensures



that the strength of each node, $s_i = \sum_j w_{ij}$, equals the number of articles where the PACS code has been listed³ (excluding articles with single PACS codes that are not part of the network).

Spearman rank correlation, and dissimilarity coefficient. If r_i^t, \dots, r_i^N represent the degree (strength) ranks of the PACS codes for year t , then the Spearman rank correlation C^S between the years t and t' is defined as

$$C_{t,t'}^S = \frac{\sum_i [r_i^t - \langle r^t \rangle] [r_i^{t'} - \langle r^{t'} \rangle]}{\sqrt{\sum_i [r_i^t - \langle r^t \rangle]^2 \sum_i [r_i^{t'} - \langle r^{t'} \rangle]^2}}, \quad (1)$$

where $\langle \dots \rangle$ represents the average over all nodes. From C^S we calculate the dissimilarity coefficient $\zeta \equiv 1 - (C^S)^2$, where $\zeta \in [0, 1]$, with low values indicating that the rank of the individual nodes remain invariant over time¹².

Weighted overlap. In a unweighted network, the overlap is used to determine the similarity in the neighborhood of two nodes²¹. However, if the network is weighted and the link weight distribution is heterogeneous, one should put more significance on links having large weights. In order to do this we define the weighted version of the neighborhood overlap O_{ij}^w between nodes i and j as

$$O_{ij}^w = \frac{W_{ij}}{s_i + s_j - 2 \times w_{ij} - W_{ij}} \quad (2)$$

where $W_{ij} = \sum_{k \in \Lambda_i \cap \Lambda_j} (w_{ik} + w_{jk}) / 2$ and Λ_i denotes the neighborhood of node i . Thus, $O_{ij}^w = 0$ if the two nodes i and j have no common neighbors, and $O_{ij}^w = 1$ if all of their strength is associated with links to common neighbors (except for the weight of the link joining i and j , if any).

k -core analysis. We start by recursively removing nodes that have a single link until no such nodes remain in the network. These nodes form the 1-shell of the network (k^s -shell index $k^s = 1$). Similarly, by recursively removing all nodes with degree 2, we get the 2-shell. We continue increasing k until all nodes in the network have been assigned to one of the shells. The union of all the shells with index greater than or equal to k^s is called the k^s -core of the network, and the union of all shells with index smaller or equal to k^s is the k^s -crust of the network (see also Supplementary Information).

1. Popper, K. R. *The Logic of Scientific Discovery* (Basic Books, New York, USA, 1959).
2. Kuhn, T. *The Structure of Scientific Revolutions* (University of Chicago press, Chicago, 1962).
3. Newman, M. E. J. Scientific collaboration networks. ii. shortest paths, weighted networks, and centrality. *Phys. Rev. E* **64**, 016132 (2001).
4. Redner, S. How popular is your paper? an empirical study of the citation distribution. *Eur. Phys. J. B* **4**, 131–134 (1998).
5. Pan, R. K. & Saramaki, J. The strength of strong ties in scientific collaboration networks. *Europhys. Lett.* **97**, 18007 (2012).
6. Redner, S. Citation statistics from 110 years of physical review. *Phys. Today* **58**, 49–54 (2005).

7. Gautreau, A., Barrat, A. & Barthélemy, M. Microdynamics in stationary complex networks. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 8847 (2009).
8. <http://publish.aps.org/> (accessed on 01-02-2011).
9. <http://www.aip.org/pacs/> (accessed on 01-02-2011).
10. Radicchi, F. & Castellano, C. Rescaling citations of publications in physics. *Phys. Rev. E* **83**, 046116 (2011).
11. Massey Jr, F. The Kolmogorov-Smirnov test for goodness of fit. *J. Am. Stat. Assoc.* **46**, 68–78 (1951).
12. Kossinets, G. & Watts, D. J. Empirical analysis of an evolving social network. *Science* **311**, 88–90 (2006).
13. Granovetter, M. The strength of weak ties. *Am. J. Sociol.* **78**, 1360–1380 (1973).
14. Liben-Nowell, D., Novak, J., Kumar, R., Raghavan, P. & Tomkins, A. Geographic routing in social networks. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 11623–11628 (2005).
15. Onnela, J.-P., Chakraborti, A., Kaski, K., Kertész, J. & Kanto, A. Asset trees and asset graphs in financial markets. *Phys. Scripta* **T106**, 48–54 (2003).
16. Bollobás, B. *Graph in Theory and Combinatorics: Proceedings of the Cambridge Combinatorial Conference in Honor of P. Erdős* (Academic, 1984), pp. 35–57.
17. Seidman, S. Network structure and minimum degree. *Social Networks* **5**, 269–287, pp. (1983).
18. Carmi, S., Havlin, S., Kirkpatrick, S., Shavitt, Y. & Shir, E. A model of internet topology using k-shell decomposition. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 11150–11154 (2007).
19. Kitsak, M. *et al.* Identification of influential spreaders in complex networks. *Nature Phys.* **6**, 888–893 (2010).
20. Rosvall, M. & Bergstrom, C. T. Mapping change in large networks. *PLoS ONE* **5**, e8694 (2010).
21. Onnela, J.-P. *et al.* Structure and tie strengths in mobile communication networks. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 7332 (2007).

Acknowledgments

Financial support from EU's 7th Framework Program's FET-Open to ICTeCollective project no. 238597 and by the Academy of Finland, the Finnish Center of Excellence program 2006–2011, project no. 129670 are gratefully acknowledged. We would like to thank S Sanyal and A Basu for helpful discussions.

Author contributions

All authors designed the research and participated in the writing of the manuscript. RKP collected the data, analysed the data and performed the research.

Additional information

Supplementary information accompanies this paper at <http://www.nature.com/scientificreports>

Competing financial interests: The authors declare no competing financial interests.

License: This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 Unported License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-sa/3.0/>

How to cite this article: Pan, R.K., Sinha, S., Kaski, K. & Saramäki, J. The evolution of interdisciplinarity in physics research. *Sci. Rep.* **2**, 551; DOI:10.1038/srep00551 (2012).