nature portfolio

https://doi.org/10.1038/s41562-021-01084-x

Supplementary information

Fresh teams are associated with original and multidisciplinary research

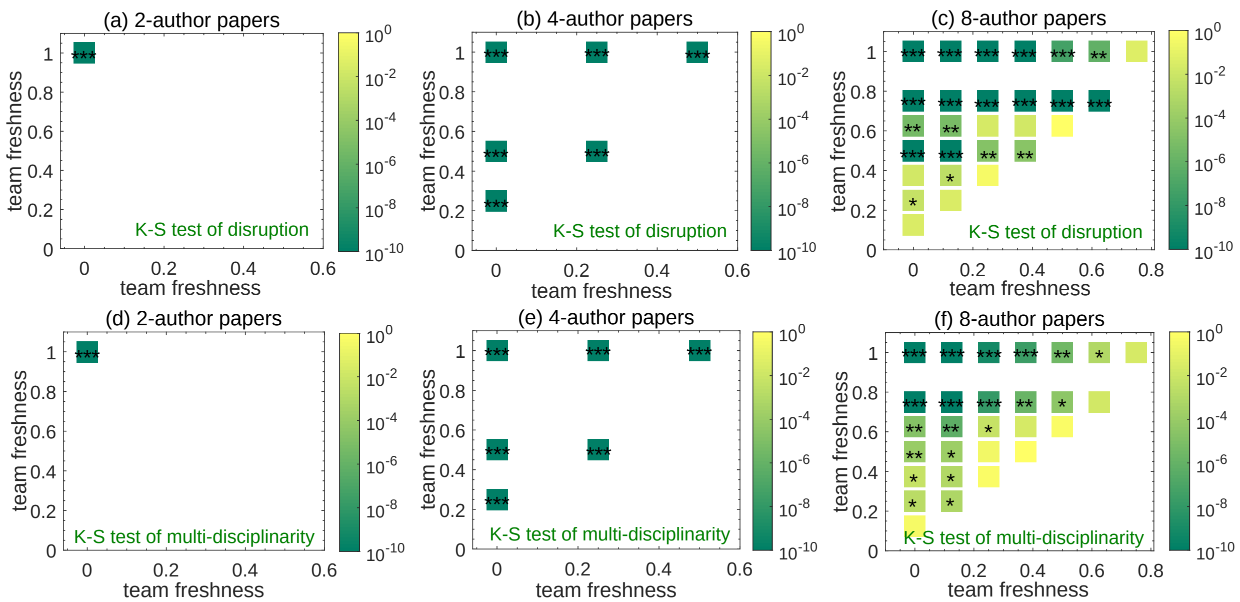
In the format provided by the authors and unedited

Supplementary Information

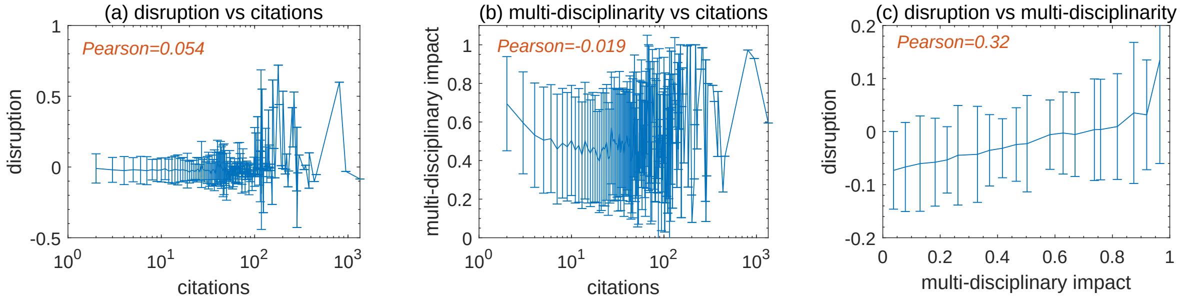
Fresh teams are associated with original and multi-disciplinary research

An Zeng, Ying Fan, Zengru Di, Yougui Wang, Shlomo Havlin

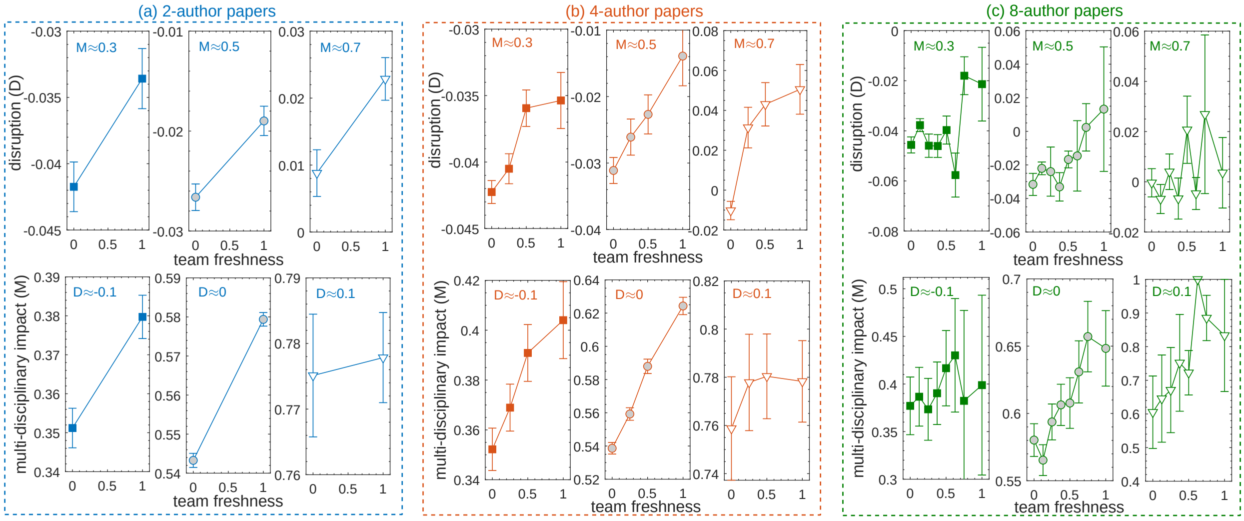
Supplementary Figures



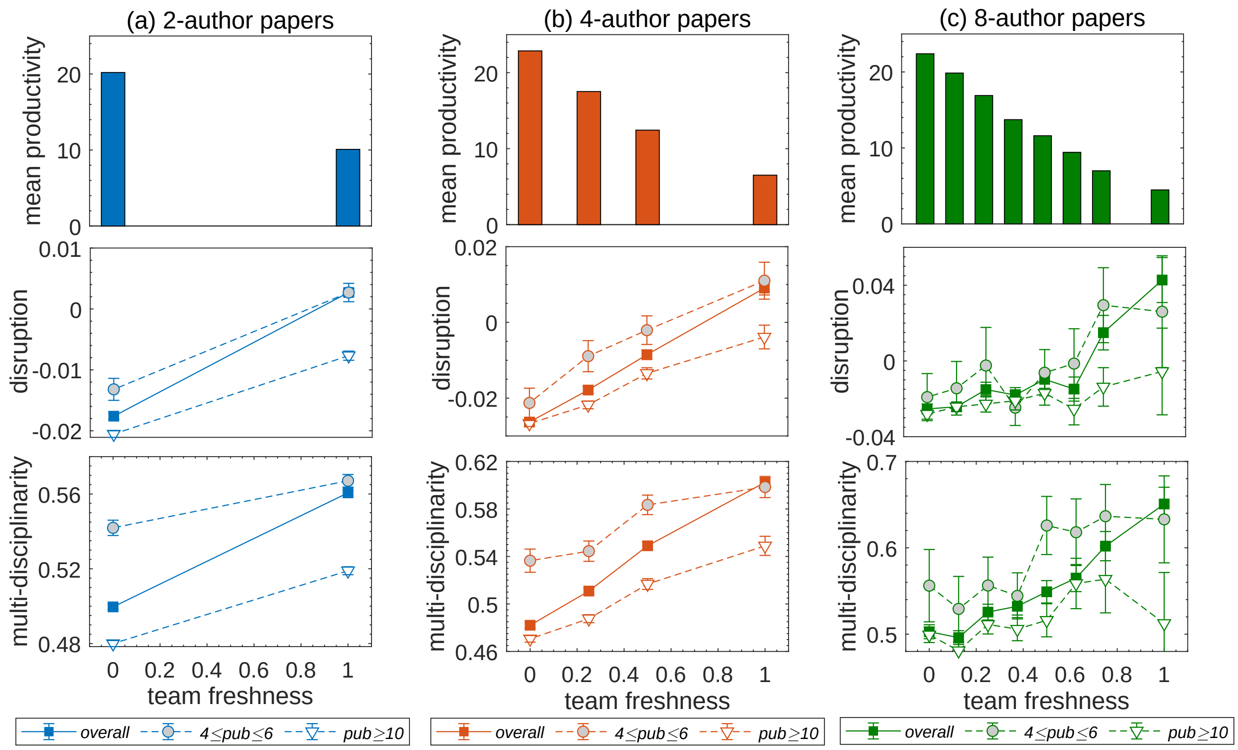
Supplementary Figure 1: Kolmogorov-Smirnov test of the significance of differences (in disruption D or multi-disciplinarity M) between papers with different team freshness. (a)-(c) We test the differences between the distributions of disruption for each pair of team freshness from zero to one using the two-tailed Kolmogorov-Smirnov test. The results for (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers are shown. (d)-(f) We also test differences between distributions of multi-disciplinarity for each pair of team freshness from zero to one using the Kolmogorov-Smirnov test. The results for (d) 2-author papers, (e) 4-author papers, and (f) 8-author papers are shown. For more strict test of the distribution differences, we take the actual values of disruption and multi-disciplinarity instead of averaged values from bootstrap for the Kolmogorov-Smirnov test. Asterisks under the numbers indicate p values: $*$ for $p \leq 0.1$, $**$ for $p \leq 0.01$, $***$ for $p \leq 0.001$. Almost all pairs of tested distributions significantly differ from one another.



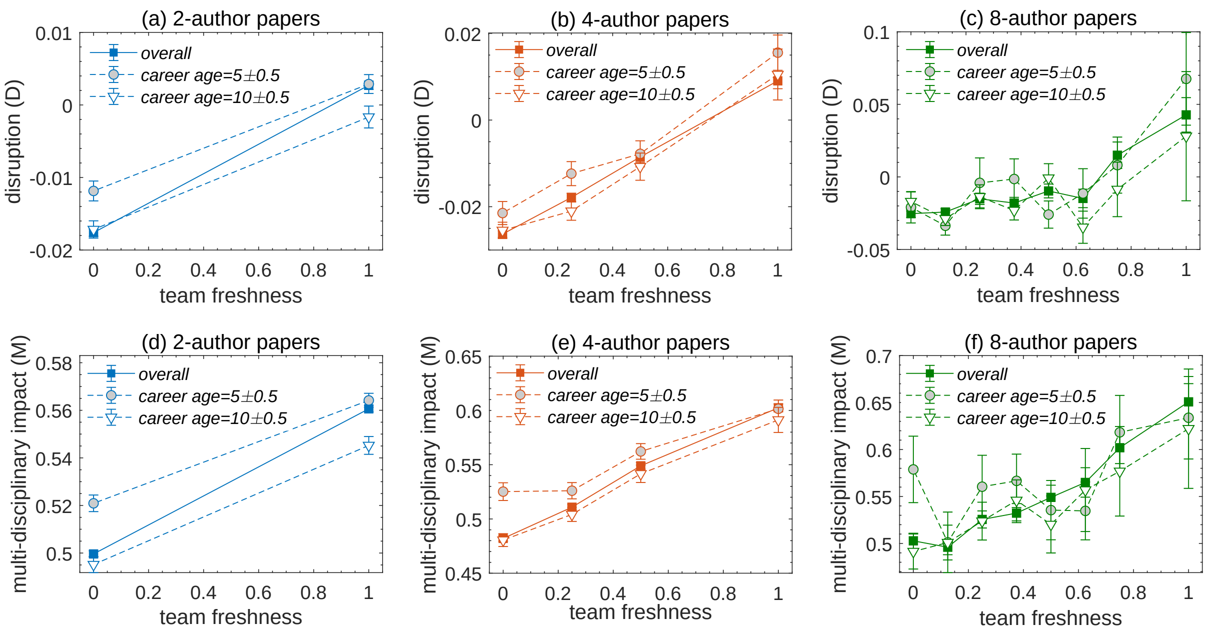
Supplementary Figure 2: Correlation between disruption and multi-disciplinary impact. (a) The average disruption of papers with different citations. The trend indicates that disruption (representing originality) is only very weakly correlated with citations. This finding is supported by directly performing the two-tailed Pearson correlation test between disruption and citations, which results in the correlation coefficient with 95% confidence interval as 0.054 ± 0.003 ($p < 0.001$). (b) The average multi-disciplinary impact for papers with different citations. Similar to disruption, only very weak correlation is observed. This is confirmed by the Pearson correlation coefficient with 95% confidence interval as -0.019 ± 0.004 ($p < 0.001$). (c) The average disruption for papers of different multi-disciplinary impact. An increasing trend is observed, indicating that papers with higher multi-disciplinary impact generally have higher disruption. However, the Pearson correlation coefficient with 95% confidence interval is only 0.320 ± 0.003 ($p < 0.001$), suggesting that these two measures are still different from each other.



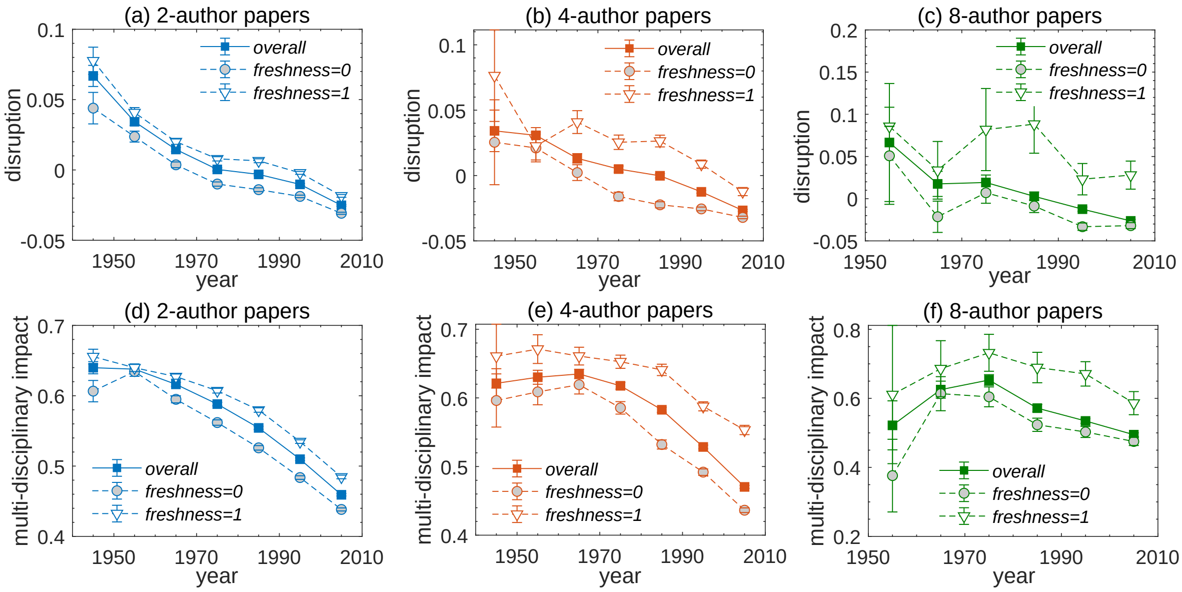
Supplementary Figure 3: Correlations when fixing disruption or multi-disciplinarity. As disruption D is to some degree correlated with multi-disciplinary impact M, a crucial question is whether the increasing trends of both indexes with team freshness measure the same feature. To address this question, we study the dependence of disruption on team freshness when controlling multi-disciplinarity, M. Specifically, we fix (top figures) M as 0.3 ± 0.01, 0.5 ± 0.01 and 0.7 ± 0.01 in (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers, respectively. The results suggest that the increasing trend of disruption with team freshness occurs even for fixed multi-disciplinarity. The increasing trend is supported by directly performing the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.051 ± 0.035 ($p = 0.005$) for $M \approx 0.3$, 0.034 ± 0.017 ($p < 0.001$) for $M \approx 0.5$, 0.070 ± 0.045 ($p = 0.003$) for $M \approx 0.7$ in 2-author papers; 0.037 ± 0.049 ($p = 0.150$) for $M \approx 0.3$, 0.057 ± 0.026 ($p < 0.001$) for $M \approx 0.5$, 0.163 ± 0.069 ($p < 0.001$) for $M \approx 0.7$ in 4-author papers; 0.008 ± 0.146 ($p = 0.909$) for $M \approx 0.3$, 0.074 ± 0.067 ($p = 0.031$) for $M \approx 0.5$, 0.057 ± 0.077 ($p = 0.153$) for $M \approx 0.7$ in 8-author papers. In addition, we study (bottom figures) the dependence of multi-disciplinarity on team freshness while controlling disruption. We fix D as -0.1 ± 0.01, 0 ± 0.01 and 0.1 ± 0.01 in (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers, respectively. The results suggest that the increasing trend of multi-disciplinarity with team freshness occurs even for fixed disruption. The increasing trend is also supported by the two-tailed Pearson correlation test between M and team freshness (coefficients with 95% confidence intervals and p-values): 0.058 ± 0.029 ($p < 0.001$) for $D \approx -0.1$, 0.065 ± 0.008 ($p < 0.001$) for $D \approx 0$, 0.006 ± 0.048 ($p = 0.814$) for $D \approx 0.1$ in 2-author papers; 0.075 ± 0.041 ($p < 0.001$) for $D \approx -0.1$, 0.107 ± 0.014 ($p < 0.001$) for $D \approx 0$, 0.023 ± 0.079 ($p = 0.572$) for $D \approx 0.1$ in 4-author papers; 0.041 ± 0.107 ($p = 0.459$) for $D \approx -0.1$, 0.086 ± 0.040 ($p < 0.001$) for $D \approx 0$, 0.115 ± 0.201 ($p = 0.285$) for $D \approx 0.1$ in 8-author papers.



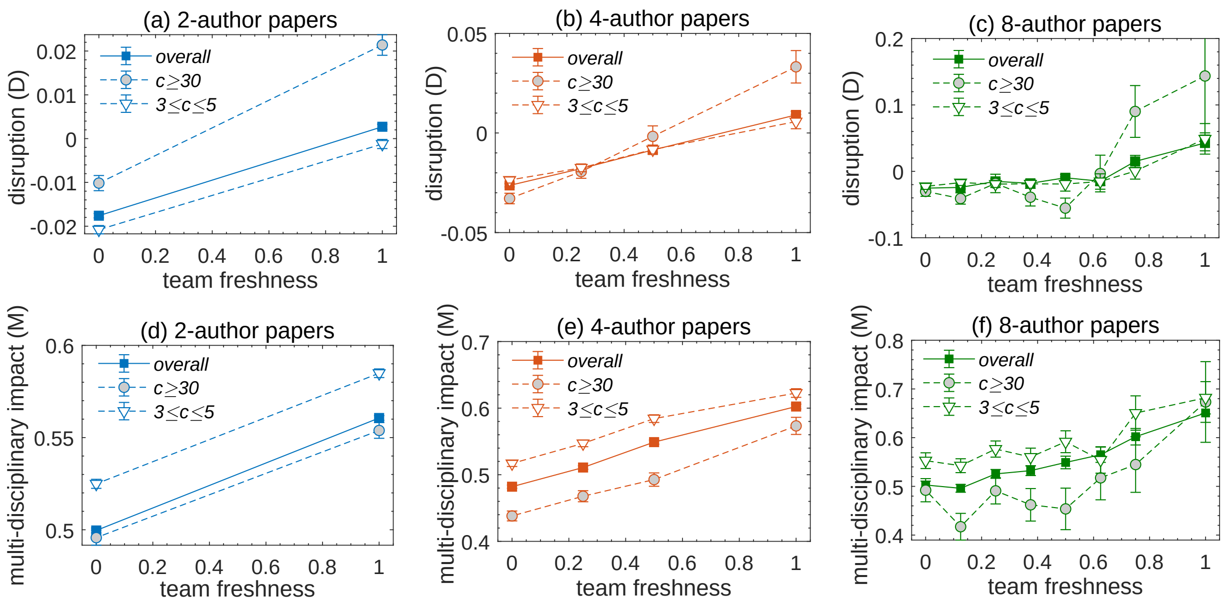
Supplementary Figure 4: Freshness in teams with similar team member productivity. In the top three panels, we show the mean prior productivity of team members in teams with size (a) 2, (b) 4 and (c) 8, respectively. The prior productivity of a team member is defined as the number of his/her papers before he/she publishes the focal paper. The results suggest that scientists in fresh teams tend to have lower prior productivity than those in old teams. In the second row, we show the increase of disruption with freshness for papers of team size 2, 4 and 8. To support the trend, we also plot the results when controlling the team member productivity. Accordingly, we add results of teams with mean team member productivity between 4 and 6, and teams with mean team member productivity larger than 10. In both cases the increased trend of disruption with freshness are still present. In the bottom three panels, we show the increase of multi-disciplinary impact with freshness in papers of team size 2, 4 and 8. Again, these results support the trend when controlling team member productivity. We further perform the two-tailed Pearson correlation test between D and team freshness (correlations with 95% confidence intervals and p-values): 0.057 ± 0.016 ($p < 0.001$) for $4 \leq pub \leq 6$, 0.061 ± 0.008 ($p < 0.001$) for $pub \geq 10$ in 2-author papers; 0.075 ± 0.028 ($p < 0.001$) for $4 \leq pub \leq 6$, 0.071 ± 0.012 ($p < 0.001$) for $pub \geq 10$ in 4-author papers; 0.096 ± 0.080 ($p = 0.020$) for $4 \leq pub \leq 6$, 0.034 ± 0.031 ($p = 0.033$) for $pub \geq 10$ in 8-author papers. The results for the correlation test between M and team freshness are: 0.042 ± 0.017 ($p < 0.001$) for $4 \leq pub \leq 6$, 0.066 ± 0.008 ($p < 0.001$) for $pub \geq 10$ in 2-author papers; 0.082 ± 0.029 ($p < 0.001$) for $4 \leq pub \leq 6$, 0.079 ± 0.013 ($p < 0.001$) for $pub \geq 10$ in 4-author papers; 0.118 ± 0.083 ($p = 0.007$) for $4 \leq pub \leq 6$, 0.044 ± 0.033 ($p = 0.009$) for $pub \geq 10$ in 8-author papers.



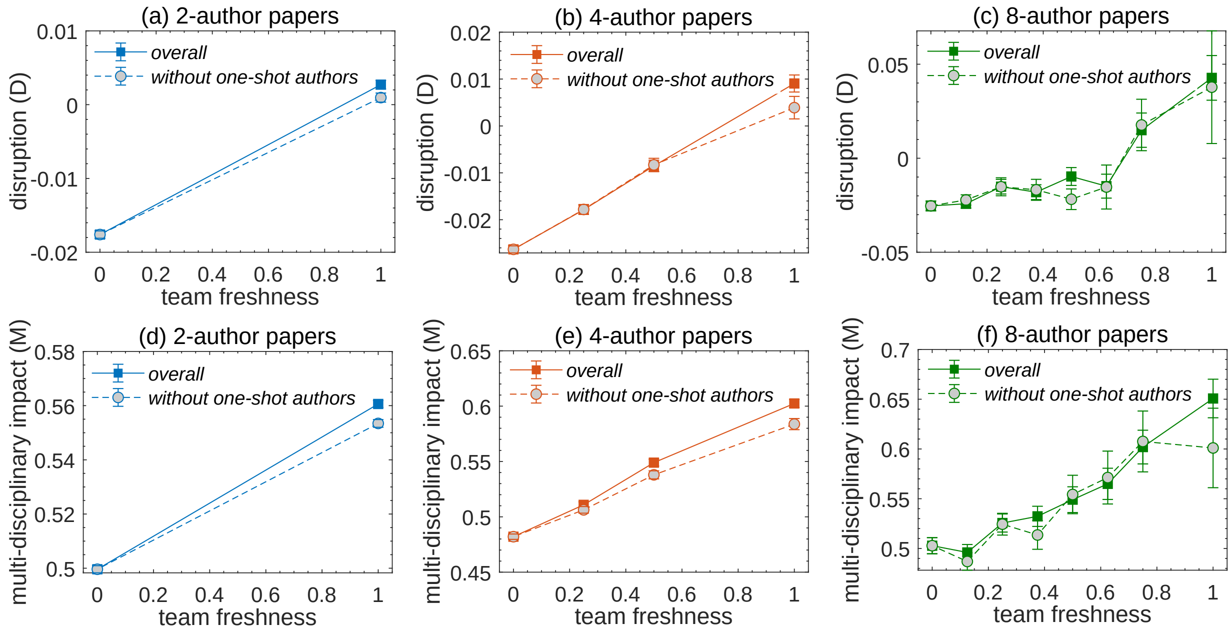
Supplementary Figure 5: Freshness in teams with similar mean career age of team members. The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. We fix the career freshness by taking the teams of mean career age of team members in [4.5, 5.5] and [9.5, 10.5] years, respectively. After controlling the career freshness, one can still observe a clear positive correlation between team freshness and disruption/multidisciplinary impact. We further perform the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.061 ± 0.015 ($p < 0.001$) for career age≈5, 0.067 ± 0.016 ($p < 0.001$) for career age≈10 in 2-author papers; 0.057 ± 0.030 ($p < 0.001$) for career age≈5, 0.097 ± 0.023 ($p < 0.001$) for career age≈10 in 4-author papers; 0.130 ± 0.108 ($p = 0.022$) for career age≈5, 0.051 ± 0.058 ($p = 0.088$) for career age≈10 in 8-author papers. The results for the correlation test between M and team freshness are: 0.067 ± 0.116 ($p = 0.270$) for career age≈5, 0.095 ± 0.063 ($p = 0.003$) for career age≈10 in 2-author papers; 0.112 ± 0.032 ($p < 0.001$) for career age≈5, 0.122 ± 0.024 ($p < 0.001$) for career age≈10 in 4-author papers; 0.139 ± 0.112 ($p = 0.019$) for career age≈5, 0.092 ± 0.061 ($p = 0.003$) for career age≈10 in 8-author papers.



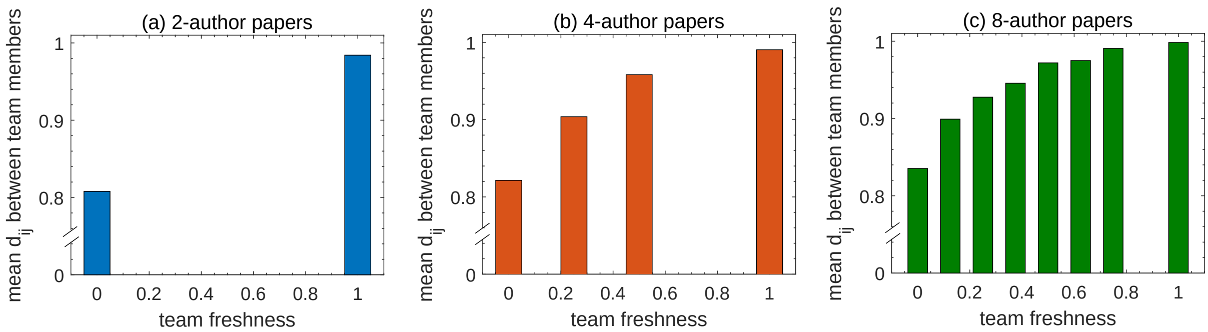
Supplementary Figure 6: **Time evolution of originality and multi-disciplinary impact of papers.** The time evolution of the mean disruption D of (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers. The time evolution of the multi-disciplinary impact M of (d) 2-author papers, (e) 4-author papers, and (f) 8-author papers. The results suggest that both indexes decrease (at least in the last 50 years) with time, which is also observed in ref. [1]. The decreasing trend is supported by performing the Pearson correlation test between D and year (coefficients with 95% confidence intervals and p-values): -0.14 ± 0.006 ($p < 0.001$) for 2-author papers; -0.123 ± 0.009 ($p < 0.001$) for 4-author papers; -0.139 ± 0.024 ($p < 0.001$) for 8-author papers. The results for the correlation test between M and year are: -0.212 ± 0.006 ($p < 0.001$) for 2-author papers; -0.189 ± 0.009 ($p < 0.001$) for 4-author papers; -0.155 ± 0.026 ($p < 0.001$) for 8-author papers. For comparison, also the papers published by old teams (freshness=0) and fresh teams (freshness=1) are shown in each panel, showing that fresh teams systematically have higher originality and multi-disciplinarity than old teams.



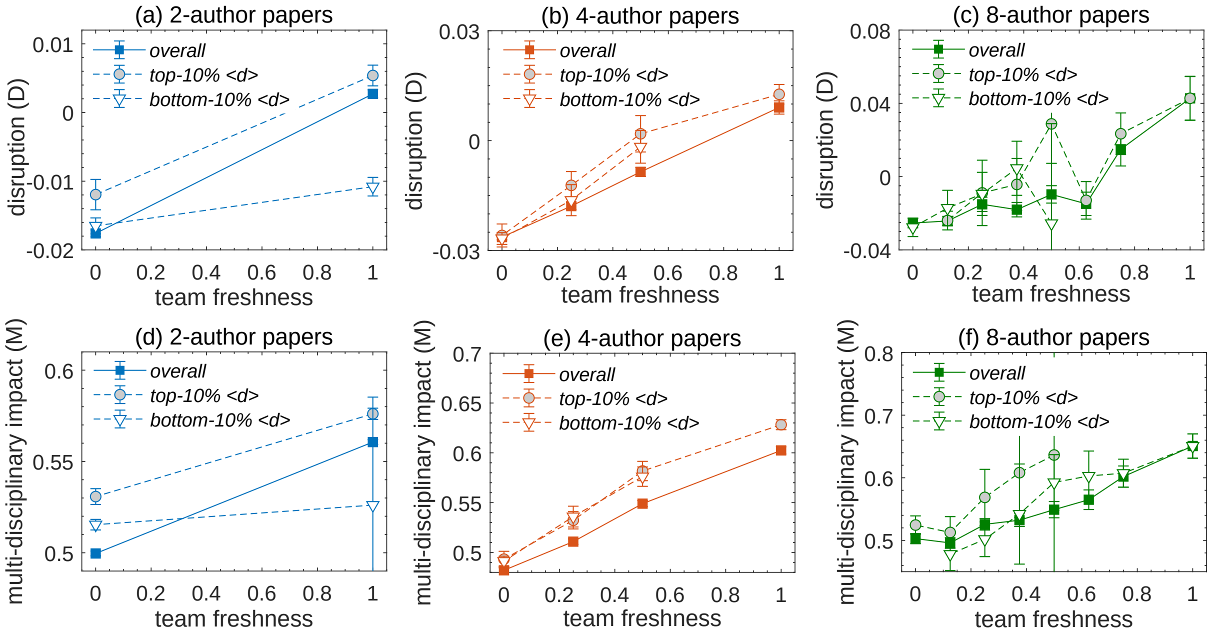
Supplementary Figure 7: Robustness test by considering only highly cited and less cited papers. The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. Here, we control the number of citations c of papers. We consider a group of highly cited papers (with at least 30 citations) and another group of less cited papers (citations between 3 and 5). The results suggest that team freshness is positively correlated with disruption/multi-disciplinarity in both highly and lowly cited papers. We further perform the Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.109 ± 0.019 ($p < 0.001$) for $c \geq 30$, 0.084 ± 0.012 ($p < 0.001$) for $3 \leq c \leq 5$ in 2-author papers; 0.165 ± 0.031 ($p < 0.001$) for $c \geq 30$, 0.093 ± 0.019 ($p < 0.001$) for $3 \leq c \leq 5$ in 4-author papers; 0.180 ± 0.081 ($p < 0.001$) for $c \geq 30$, 0.116 ± 0.051 ($p < 0.001$) for $3 \leq c \leq 5$ in 8-author papers. The results for the correlation test between M and team freshness are: 0.098 ± 0.019 ($p < 0.001$) for $c \geq 30$, 0.107 ± 0.012 ($p < 0.001$) for $3 \leq c \leq 5$ in 2-author papers; 0.150 ± 0.031 ($p < 0.001$) for $c \geq 30$, 0.138 ± 0.018 ($p < 0.001$) for $3 \leq c \leq 5$ in 4-author papers; 0.083 ± 0.082 ($p = 0.054$) for $c \geq 30$, 0.108 ± 0.050 ($p < 0.001$) for $3 \leq c \leq 5$ in 8-author papers.



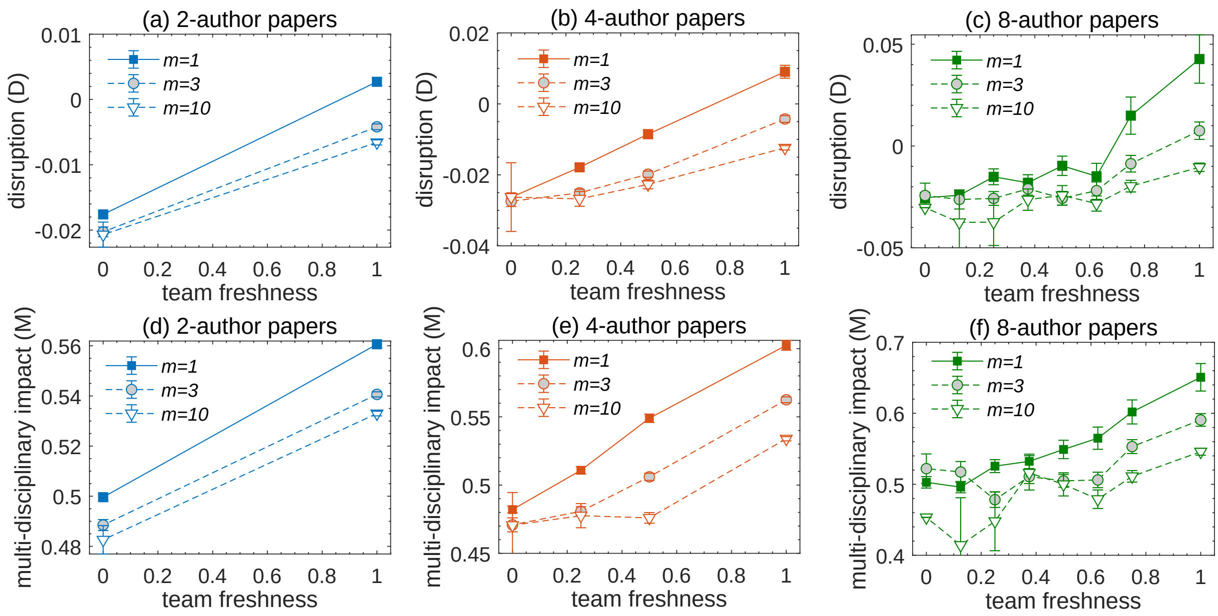
Supplementary Figure 8: **Robustness test by removing one-shot authors.** The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. There are around 43% authors in APS are one-shot authors (i.e. authors publish only one paper in APS). As these one-shot authors might inflate team freshness, we must investigate the relation between team freshness and disruption/multi-disciplinarity in the APS data where all one-shot authors are removed. We find interestingly that these one-shot authors represent only 15.7% papers in APS. This is because multiple one-shot authors often cluster in one paper. We carry out the analysis in the papers without any one-shot authors. We observe that the curves after removing one-shot authors overlap well with those obtained from the original data, indicating that our findings are robust and not caused by the one-shot authors. The results for the two-tailed Pearson correlation test between D and team freshness without one-shot authors are (coefficients with 95% confidence intervals and p-values): 0.074 ± 0.006 ($p < 0.001$) for 2-author papers; 0.091 ± 0.010 ($p < 0.001$) for 4-author papers; 0.072 ± 0.031 ($p < 0.001$) for 8-author papers. The results for the correlation test between M and team freshness without one-shot authors are: 0.093 ± 0.006 ($p < 0.001$) for 2-author papers; 0.114 ± 0.011 ($p < 0.001$) for 4-author papers; 0.085 ± 0.032 ($p < 0.001$) for 8-author papers.



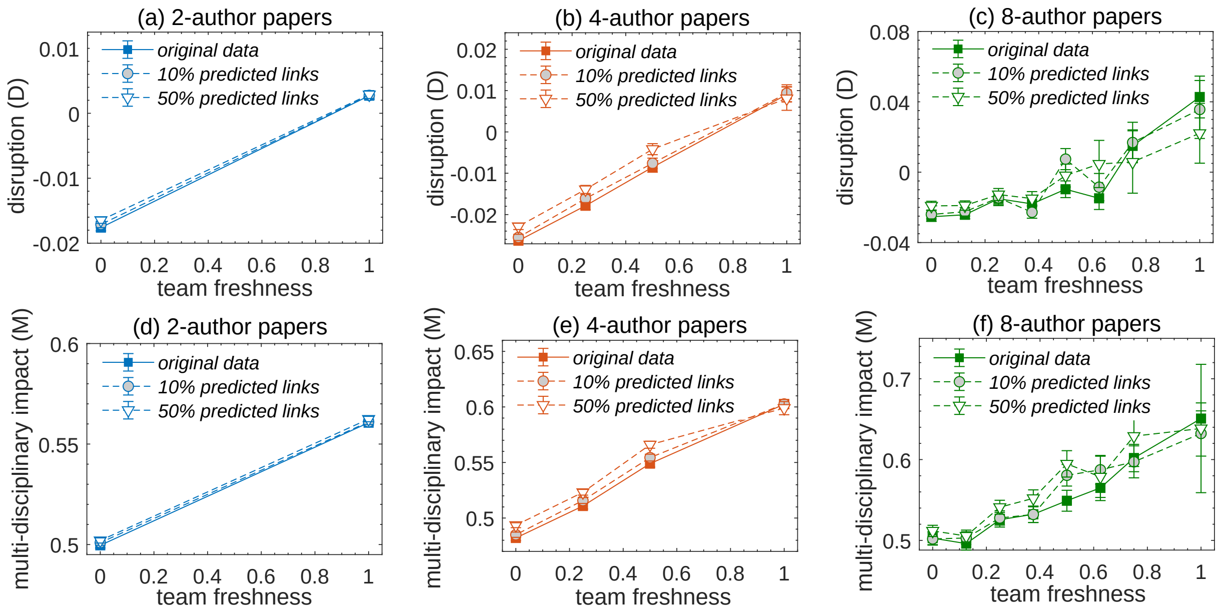
Supplementary Figure 9: **Relation between mean scientific distance of teams and team freshness.** The mean distance d_{ij} between team members in teams with different freshness for (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers. We quantify the distance d_{ij} between scientist i and scientist j in scientific space by the dis-similarity of them in their research interests. For each scientist i, we construct a set Γ_i recording all the references in his/her papers, representing the research literature he/she is interested in. The distance d_{ij} between scientist i and scientist j can thus be calculated as their Jaccard dis-similarity $d_{ij} = 1 - |\Gamma_i \cap \Gamma_j| / |\Gamma_i \cup \Gamma_j|$ where $|.|$ is the size of the set. For each paper, we calculate the mean distance d_{ij} between team members by using the data before they coauthor this paper. The mean distance d_{ij} is positively correlated with team freshness, supported by the two-tailed Pearson correlation test. The correlation coefficients with 95% confidence intervals and p-values are 0.429 ± 0.005 ($p < 0.001$) for 2-author papers, 0.401 ± 0.010 ($p < 0.001$) for 4-author papers, 0.380 ± 0.033 ($p < 0.001$) for 8-author papers. This is also supported by this figure which shows that teams with higher freshness tend to have higher mean scientific distance between team members.



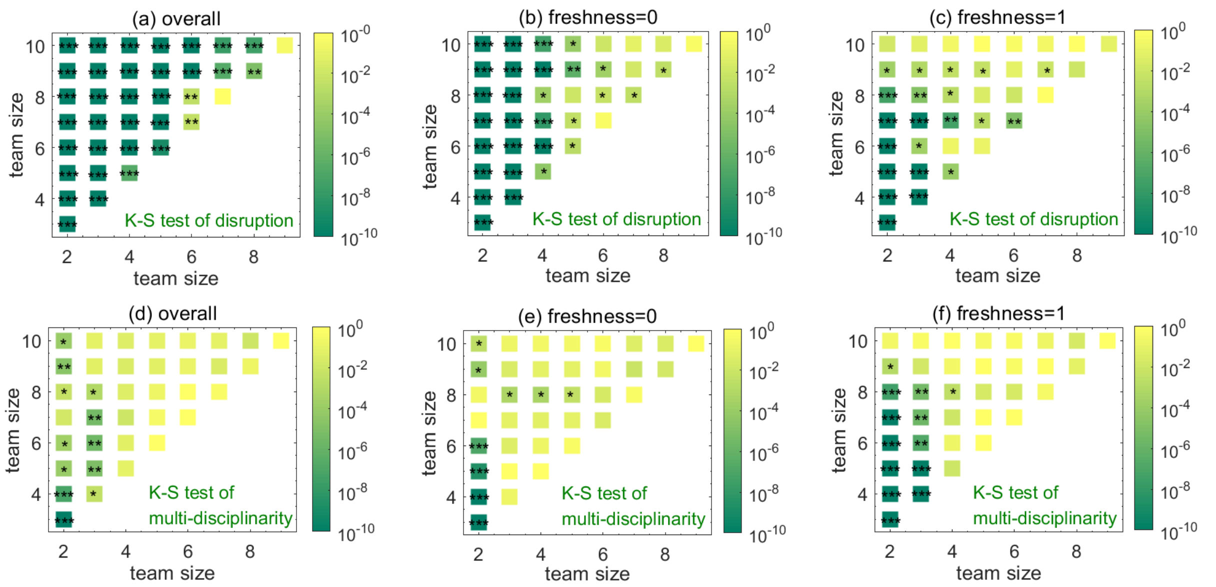
Supplementary Figure 10: **Robustness test by controlling the mean scientific distance between team members.** The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. We control here the scientific distance of team members (see the caption of Supplementary Figure 10 for the definition of scientific distance). Specifically, we analyze the 10% teams with the largest mean distance $\langle d \rangle$ between team members and the 10% teams with the smallest mean distance $\langle d \rangle$ between team members. We find that in both cases higher team freshness is associated with larger disruption and multi-disciplinary impact. In 4-author papers and 8-author papers, there are no results for large team freshness when fixing bottom 10% distance. This is because scientists who are very close in scientific space have high probability to collaborate with each other before, and thus cannot form a fully fresh team when they collaborate again. We further perform the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.056 ± 0.007 ($p < 0.001$) for top-10% $\langle d \rangle$, 0.012 ± 0.010 ($p = 0.025$) for bottom-10% $\langle d \rangle$ in 2-author papers; 0.071 ± 0.017 ($p < 0.001$) for top-10% $\langle d \rangle$, 0.041 ± 0.029 ($p = 0.006$) for bottom-10% $\langle d \rangle$ in 4-author papers; 0.087 ± 0.081 ($p = 0.040$) for top-10% $\langle d \rangle$, 0.091 ± 0.076 ($p = 0.022$) for bottom-10% $\langle d \rangle$ in 8-author papers. The results for the correlation test between M and team freshness are: 0.059 ± 0.008 ($p < 0.001$) for top-10% $\langle d \rangle$, 0.002 ± 0.019 ($p = 0.845$) for bottom-10% $\langle d \rangle$ in 2-author papers; 0.088 ± 0.018 ($p < 0.001$) for top-10% $\langle d \rangle$, 0.065 ± 0.031 ($p < 0.001$) for bottom-10% $\langle d \rangle$ in 4-author papers; 0.083 ± 0.086 ($p = 0.063$) for top-10% $\langle d \rangle$, 0.051 ± 0.080 ($p = 0.218$) for bottom-10% $\langle d \rangle$ in 8-author papers.



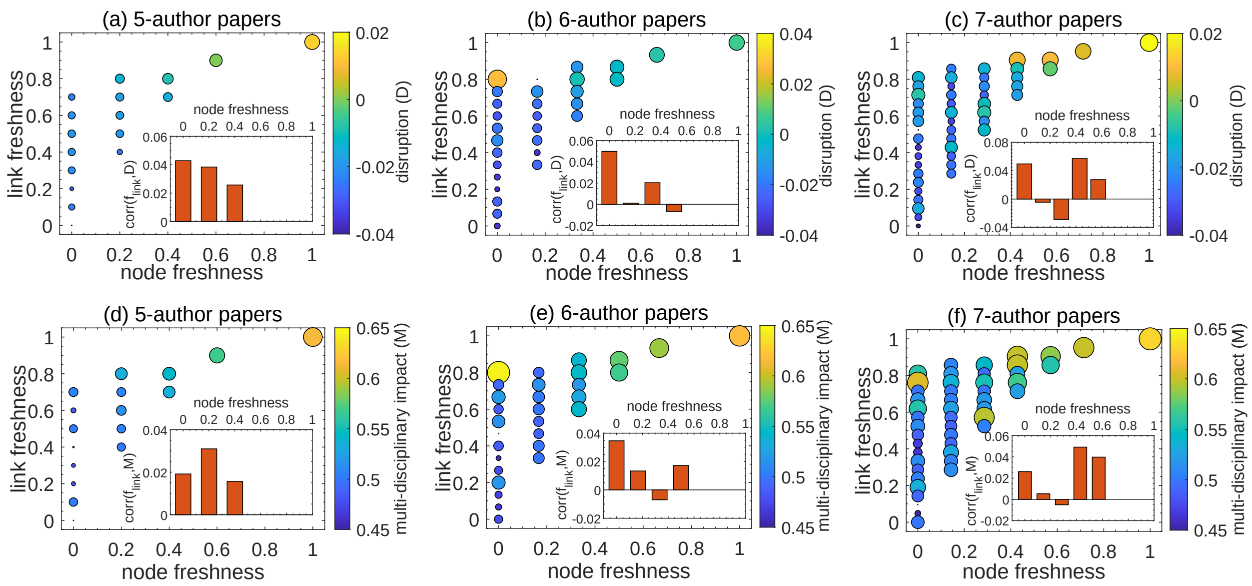
Supplementary Figure 11: **A generalization of the definition of team freshness by taking into account the strength of collaboration ties.** The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. We consider here a more general definition of team freshness in which team freshness mainly determined by strong ties. The freshness of a team of a paper is defined as the fraction of team members who have collaborated fewer than m papers with any of other team members before they coauthor this paper. In this new definition, scientists having published fewer than m coauthored papers with each other will still be regarded as a fresh team when they work together next time. We study the relation between the new team freshness and disruption/multi-disciplinarity under $m = 3$ and $m = 10$. In these settings, positive correlations can still be observed. We further perform the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.046 ± 0.006 ($p < 0.001$) for $m = 3$, 0.017 ± 0.006 ($p < 0.001$) for $m = 10$ in 2-author papers; 0.079 ± 0.009 ($p < 0.001$) for $m = 3$, 0.034 ± 0.009 ($p < 0.001$) for $m = 10$ in 4-author papers; 0.096 ± 0.025 ($p < 0.001$) for $m = 3$, 0.057 ± 0.025 ($p < 0.001$) for $m = 10$ in 8-author papers. The results for the correlation test between M and team freshness are: 0.067 ± 0.006 ($p < 0.001$) for $m = 3$, 0.027 ± 0.006 ($p < 0.001$) for $m = 10$ in 2-author papers; 0.124 ± 0.009 ($p < 0.001$) for $m = 3$, 0.069 ± 0.009 ($p < 0.001$) for $m = 10$ in 4-author papers; 0.108 ± 0.026 ($p < 0.001$) for $m = 3$, 0.078 ± 0.026 ($p < 0.001$) for $m = 10$ in 8-author papers.



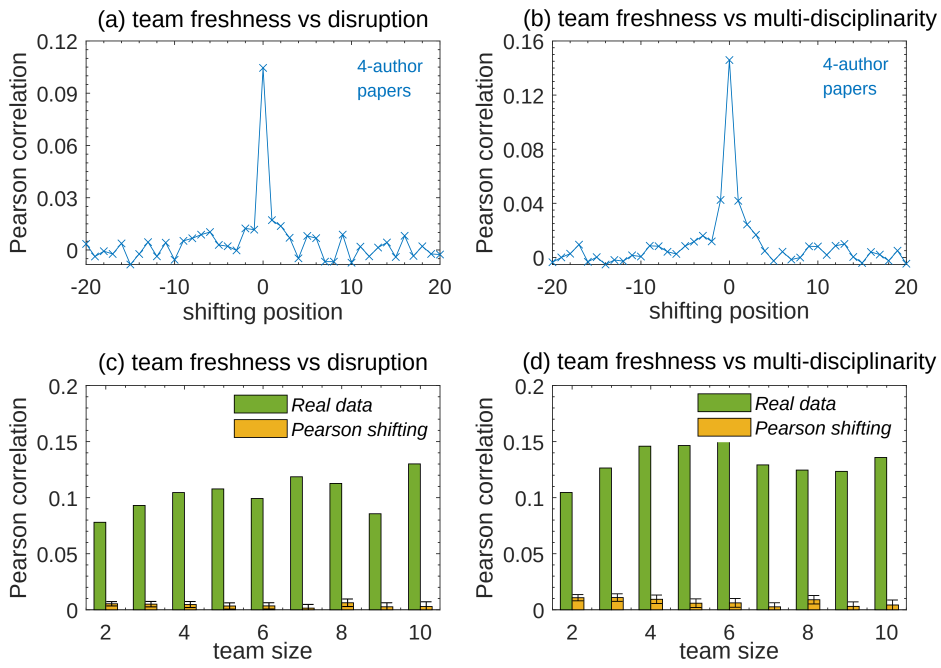
Supplementary Figure 12: **Predicting collaboration relations between scientists outside the APS data.** The dependence of the disruption D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. We use the APS data set as the training set and use the Resource Allocation link prediction algorithm [3] to predict and add missing collaboration relations to the data. The team freshness is respectively calculated in both the original APS data and the data with predicted links. Denoting E as the number of links in the original APS collaboration network, we add $10\% * E$ or $50\% * E$ most likely missing links to the collaboration network, with the likelihood estimated via the Resource Allocation method. In these cases, a team is fully fresh only if the team members have no prior collaboration in the APS data and no added missing links connecting them. The results suggest that the curves of data with predicted links overlap well with those of the original data, confirming the positive correlations between team freshness and disruption/multi-disciplinarity. We further perform the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.076 ± 0.006 ($p < 0.001$) for 10% predicted links, 0.074 ± 0.006 ($p < 0.001$) for 50% predicted links in 2-author papers; 0.097 ± 0.009 ($p < 0.001$) for 10% predicted links, 0.077 ± 0.009 ($p < 0.001$) for 50% predicted links in 4-author papers; 0.097 ± 0.025 ($p < 0.001$) for 10% predicted links, 0.047 ± 0.025 ($p < 0.001$) for 50% predicted links in 8-author papers. The results for the correlation test between M and team freshness are: 0.103 ± 0.006 ($p < 0.001$) for 10% predicted links, 0.104 ± 0.006 ($p < 0.001$) for 50% predicted links in 2-author papers; 0.136 ± 0.010 ($p < 0.001$) for 10% predicted links, 0.113 ± 0.010 ($p < 0.001$) for 50% predicted links in 4-author papers; 0.114 ± 0.026 ($p < 0.001$) for 10% predicted links, 0.084 ± 0.026 ($p < 0.001$) for 50% predicted links in 8-author papers.



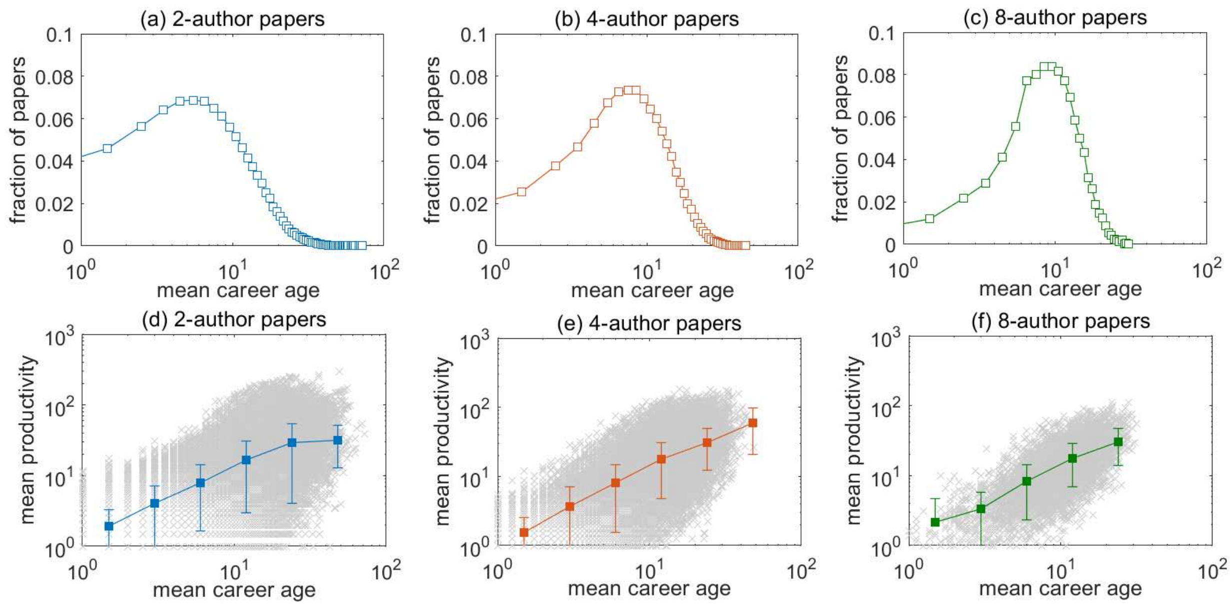
Supplementary Figure 13: **Kolmogorov-Smirnov significance test of difference of disruption** D **or multi-disciplinarity** M **between papers of different team sizes.** We test the differences between distributions of disruption for each pair of team sizes from two to ten using the two-tailed Kolmogorov-Smirnov test. We also test differences between distributions of multi-disciplinarity for each pair of team size. In addition to the overall results in (a) and (d), we show also the results for papers with freshness=0 in (b)(e) and papers with freshness=1 in (c)(f). For more strict test of the distribution differences, we take the actual values of disruption and multi-disciplinarity instead of averaged values from bootstrap for the Kolmogorov-Smirnov test. Asterisks under the numbers indicate p values: $*$ for $p \leq 0.1$, $**$ for $p \leq 0.01$, $***$ for $p \leq 0.001$. Almost all pairs of tested distributions of disruption significantly differ from one another, indicating that team size is a critical factor affecting disruption (originality). However, most of the tested distributions of multi-disciplinarity are not significantly different from each other, except when comparing team size 2 and 3 with other team sizes.



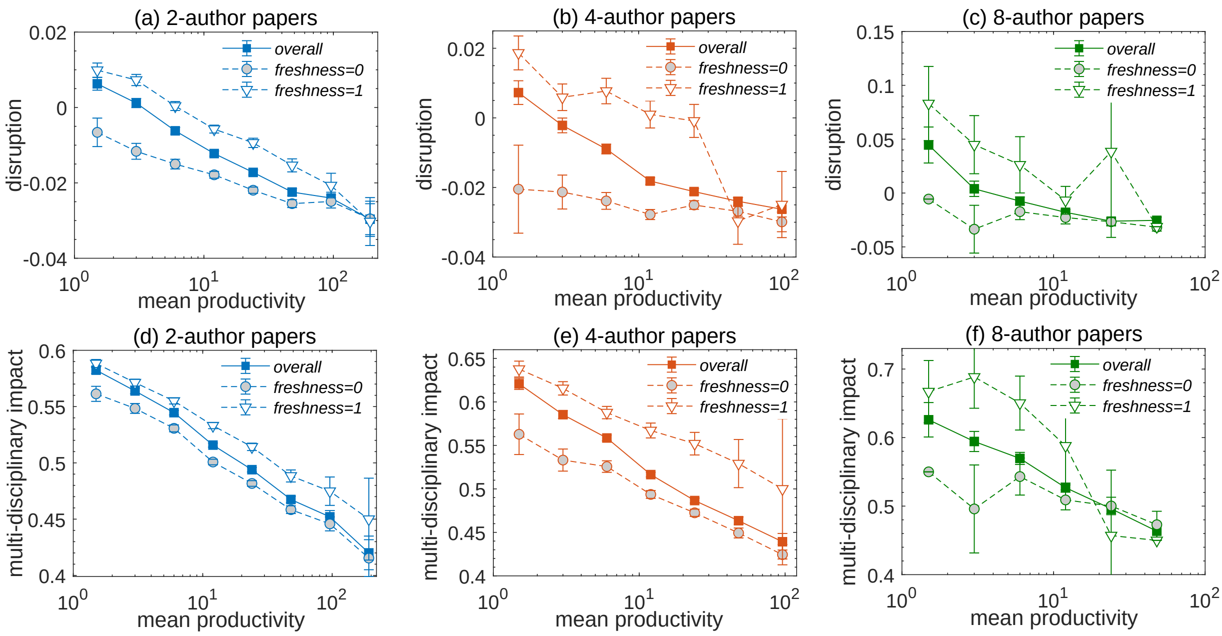
Supplementary Figure 14: **Node freshness versus link freshness for papers of 5, 6 or 7 authors.** In the manuscript, we show the dependence of disruption and multi-disciplinary impact of 8-author papers on node freshness and link freshness. In the APS data sets, the number of 8-author papers is relatively small, 6965 in total. For better statistics, we examine here also papers of 5, 6, 7 authors (There are 27895 papers of 5 authors, 16858 papers of 6 authors, 10487 papers of 7 authors). (a)-(c) The scatter plot of link freshness versus node freshness for (a) 5-author, (b) 6-author, (c) 7-author papers, with the circle size and the color representing the mean originality (disruption) of the corresponding papers. The Pearson correlation between link freshness and disruption for each node freshness is very small and shown in the insets. (d)-(f) The scatter plot of link freshness versus node freshness for (d) 5-author, (e) 6-author, (f) 7-author papers, with the circle size and the color representing the mean multi-disciplinary impact of the corresponding papers. The Pearson correlation between link freshness and multi-disciplinarity for each node freshness is very small and shown in the insets.



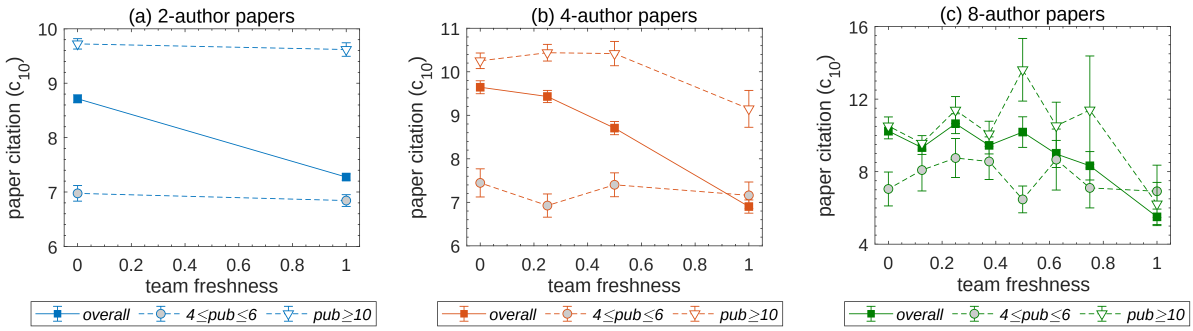
Supplementary Figure 15: **Significance of the correlation of team freshness with disruption and multi-disciplinarity.** (a) To support the significance of the correlation between team freshness and originality (disruption), see Fig. 4b, we conduct the Pearson shifting test in which the Pearson correlation is calculated after shifting all elements in the vector of team freshness by certain number of positions. We take all 4-author papers and calculate the Pearson correlation (between the vectors of freshness and originality) with shifting one vector with respect to the other by 20 positions (-20 to 20). The true correlation is without movements while each movement is like a random correlations, showing therefore the level of noise. The sharp peak at shifting zero suggests that the correlation of team freshness and originality (disruption) in the original data is indeed significant. (b) The shifting Pearson correlation between team freshness and multi-disciplinarity. A sharp peak at shifting zero can be observed. (c) The Pearson correlation between team freshness and originality (disruption) for papers with different team sizes. To support the significance of the correlation, we show the averaged Pearson correlation over different number of shifting positions from -20 to 20 (without shifting 0). The error bars represent the standard error of the mean. The results show that the Pearson correlation after shifting becomes much lower, indicating the significance of the Pearson correlation in real data. (d) The Pearson correlation between team freshness and multi-disciplinarity for papers with different team sizes. The Pearson shifting test also indicates the significance of the Pearson correlation in real data.



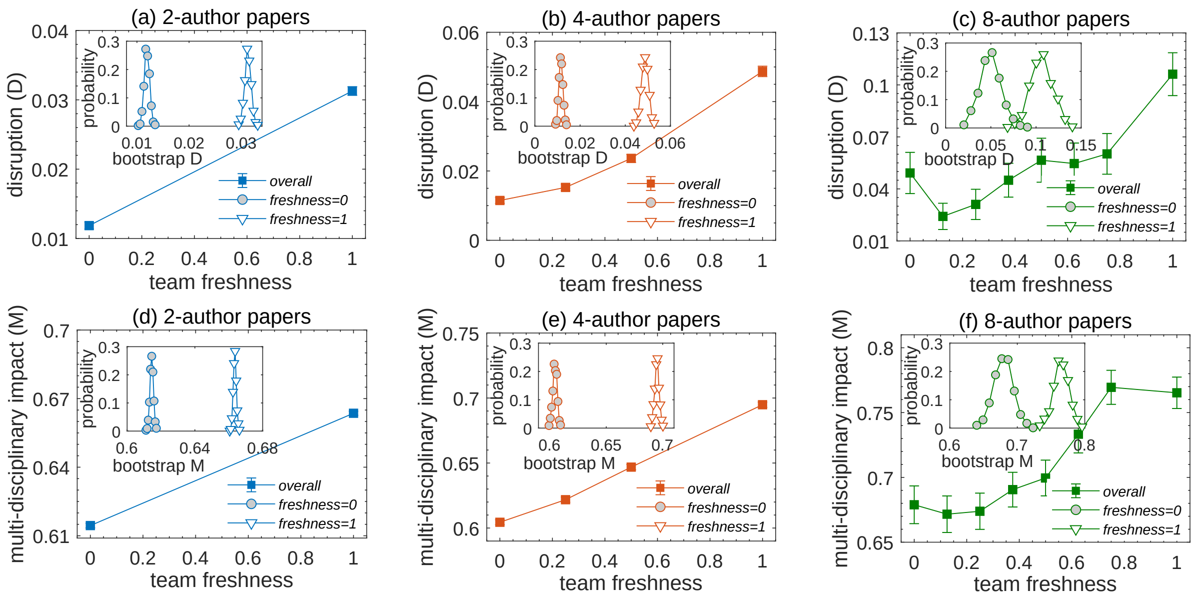
Supplementary Figure 16: **Basic statistics of mean career age of team members.** The mean career age of team members is the average career age of the team members. The mean productivity of team members is the average number of papers published by team members before the focal paper. (a)-(c) Distributions of mean career age of (a) 2-author papers, (b) 4-author papers, and (c) 8-author papers. A peak can be observed for each distribution (peaking at age=5.5 years for 2-author papers, peaking at age=8 years for 4-author papers, peaking at age=9 years for 8-author papers). (d)-(f) Scatter plot of the mean productivity versus mean career age for (d) 2-author papers, (e) 4-author papers, and (f) 8-author papers. The averaged curves indicate a clear positive correlation between mean productivity and mean career age of team members. The results for the two-tailed Pearson correlation test between mean productivity and mean career age of team members are (coefficients with 95% confidence intervals and p-values): 0.607 ± 0.003 ($p < 0.001$) for 2-author papers; 0.607 ± 0.005 ($p < 0.001$) for 4-author papers; 0.630 ± 0.014 ($p < 0.001$) for 8-author papers.



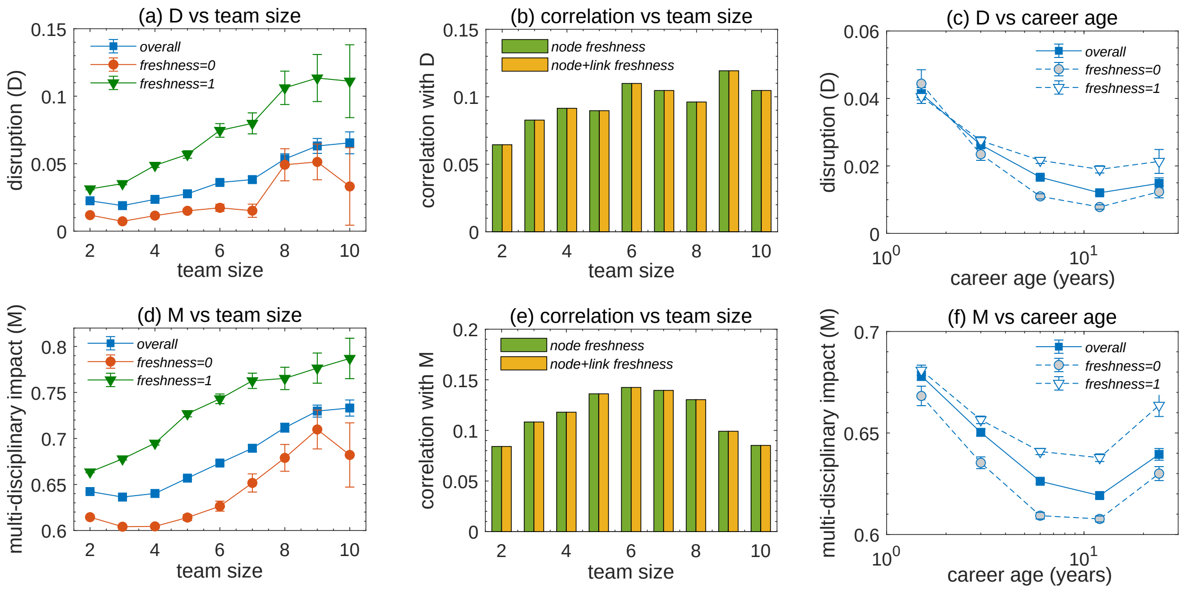
Supplementary Figure 17: Freshness based on productivity of team member's careers. The mean prior productivity of team members also represents the freshness of team members' careers. A small mean productivity indicates that team members are in the early stage of their careers, namely a fresh career. We show the dependence of the mean disruption D on team member productivity, in (a) 2-author papers, (b) 4-author papers and (c) 8-author papers. To remove the effect of team freshness, we show the dependence of disruption on team member productivity for old teams (freshness $f = 0$) and fresh team (freshness $f = 1$), respectively. The results suggest that fresher career teams are associated with more original works. The results for the two-tailed Pearson correlation test between D and mean prior productivity are (coefficients with 95% confidence intervals and p-values): -0.058 ± 0.014 ($p < 0.001$) for overall, -0.028 ± 0.008 ($p < 0.001$) for $f = 0$, -0.055 ± 0.008 ($p < 0.001$) for $f = 1$ in 2-author papers; -0.062 ± 0.009 ($p < 0.001$) for overall, -0.015 ± 0.016 ($p = 0.074$) for $f = 0$, -0.052 ± 0.023 ($p < 0.001$) for $f = 1$ in 4-author papers; -0.070 ± 0.025 ($p < 0.001$) for overall, -0.026 ± 0.054 ($p = 0.350$) for $f = 0$, -0.122 ± 0.124 ($p = 0.055$) for $f = 1$ in 8-author papers. We show also the dependence of the multi-disciplinary impact M on team member productivity, in (d) 2-author papers, (e) 4-author papers and (f) 8-author papers. The results suggest that fresher career teams are associated with works with more diverse impact. The results for the correlation test between M and mean prior productivity are: -0.118 ± 0.006 ($p < 0.001$) for overall, -0.089 ± 0.009 ($p < 0.001$) for $f = 0$, -0.107 ± 0.008 ($p < 0.001$) for $f = 1$ in 2-author papers; -0.135 ± 0.010 ($p < 0.001$) for overall, -0.091 ± 0.017 ($p < 0.001$) for $f = 0$, -0.106 ± 0.024 ($p < 0.001$) for $f = 1$ in 4-author papers; -0.125 ± 0.026 ($p < 0.001$) for overall, -0.077 ± 0.057 ($p = 0.008$) for $f = 0$, -0.150 ± 0.131 ($p = 0.025$) for $f = 1$ in 8-author papers.



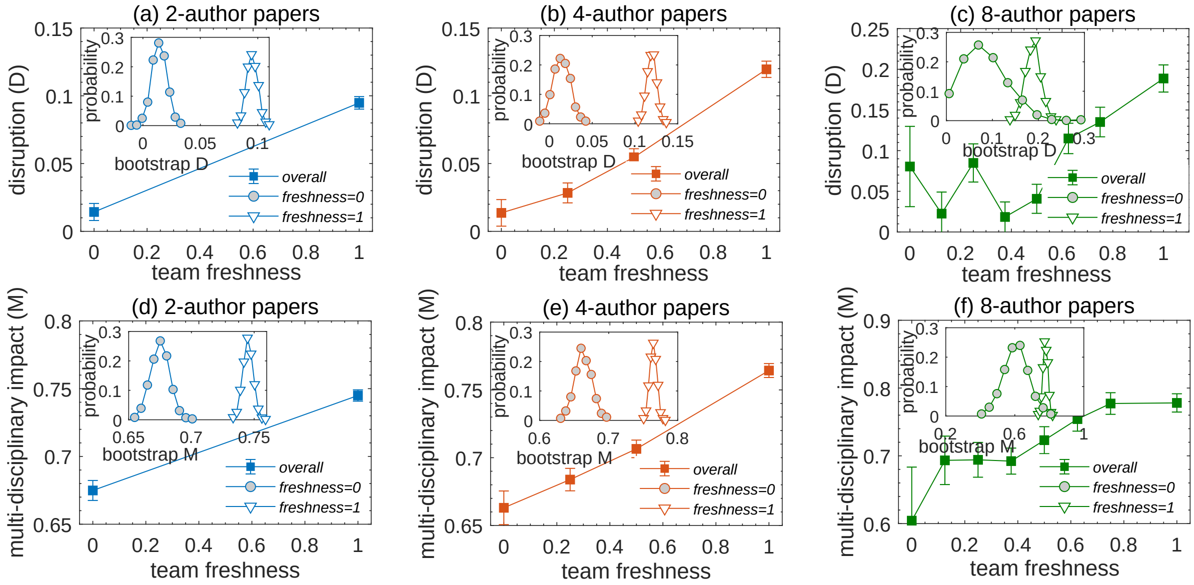
Supplementary Figure 18: **Relation between team freshness and paper citations.** For each paper, we calculate the number of citations it received during ten years after its publication, denoted as c_{10}. In this figure, we show the citation c_{10} for papers of different team freshness, in (a) 2-author papers, (b) 4-author papers, (c) 8-author papers, respectively. It is seen that citation decreases with freshness, which is consistent with the findings in ref. [2]. For comparison, we consider papers published by the teams with team members' prior productivity around 5 ($4 \leq pub \leq 6$) and higher than 10 ($pub \geq 10$). After controlling team members' productivity, one can see, however, that papers with different team freshness does not exhibit significant difference in c_{10}. We further perform the two-tailed Pearson correlation test between team freshness and citation c_{10}, and obtain (coefficients with 95% confidence intervals and p-values): -0.043 ± 0.005 ($p < 0.001$) for overall, -0.006 ± 0.015 ($p = 0.454$) for $4 \leq pub \leq 6$, -0.003 ± 0.008 ($p = 0.506$) for $pub \geq 10$ in 2-author papers; -0.056 ± 0.009 ($p < 0.001$) for overall, -0.003 ± 0.026 ($p = 0.834$) for $4 \leq pub \leq 6$, -0.009 ± 0.012 ($p = 0.141$) for $pub \geq 10$ in 4-author papers; -0.047 ± 0.024 ($p < 0.001$) for overall, -0.023 ± 0.077 ($p = 0.553$) for $4 \leq pub \leq 6$, 0.017 ± 0.030 ($p = 0.279$) for $pub \geq 10$ in 8-author papers. One can see $p > 0.1$ in all control cases, indicating no significant correlation between team freshness and citation c_{10} when fixing team members' prior productivity.



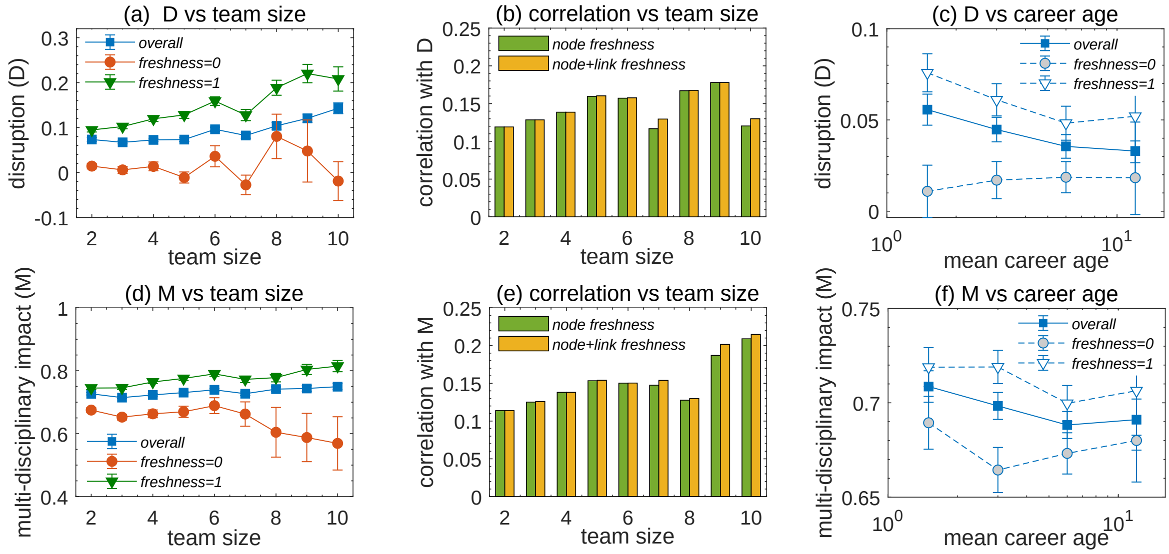
Supplementary Figure 19: Dependence of disruption (originality) and multi-disciplinarity on team freshness in computer science data. Shown are the dependence of the disruption (originality) D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. The results suggest that both originality and multi-disciplinarity significantly increase with team freshness. The insets show the distributions of bootstrap disruption or bootstrap multi-disciplinarity. A remarkable difference, i.e., high significance, can be observed between the distributions of D of papers with team freshness 0 and 1. The increasing trend is supported by performing the Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.065 ± 0.005 ($p < 0.001$) for 2-author papers; 0.092 ± 0.007 ($p < 0.001$) for 4-author papers; 0.096 ± 0.036 ($p < 0.001$) for 8-author papers. The results for the correlation test between M and team freshness are: 0.084 ± 0.005 ($p < 0.001$) for 2-author papers; 0.118 ± 0.007 ($p < 0.001$) for 4-author papers; 0.131 ± 0.034 ($p < 0.001$) for 8-author papers.



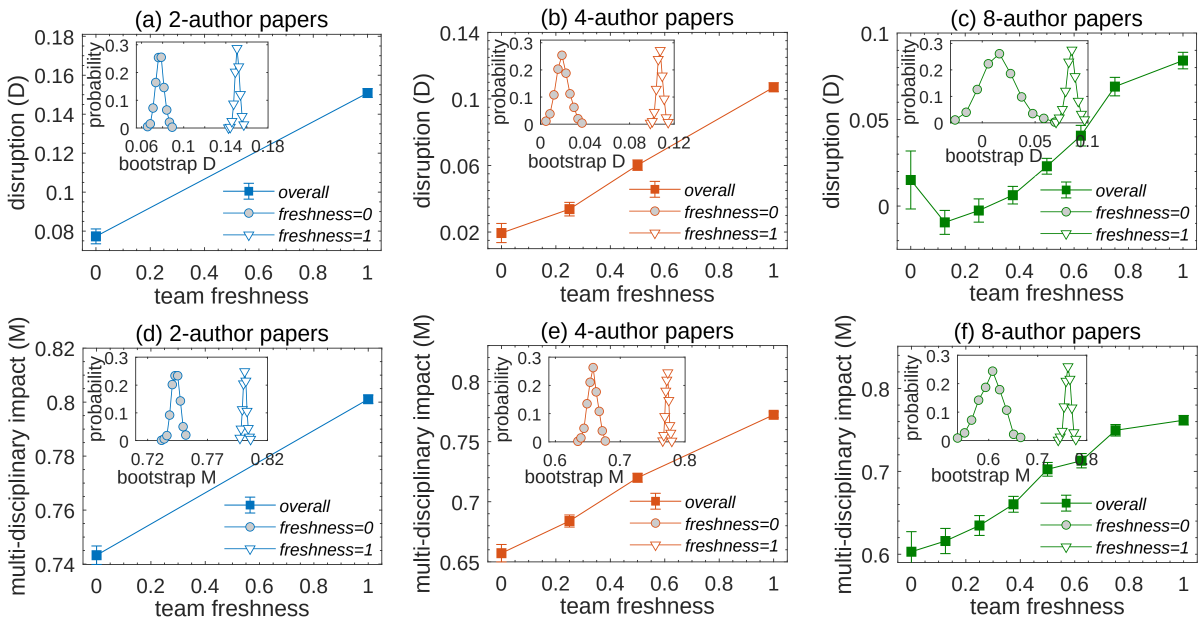
Supplementary Figure 20: **Other results of freshness in computer science data.** Plot of (a) the mean disruption (originality) D and (d) the mean multi-disciplinary impact M of papers versus different team sizes. For each team size, we also study the mean disruption D and the mean multi-disciplinary impact M of papers published by old teams (freshness=0) and fresh teams (freshness=1). The results for the two-tailed Pearson correlation test between D and team size are (coefficients with 95% confidence intervals and p-values): 0.029 ± 0.003 ($p < 0.001$) for overall; 0.010 ± 0.005 ($p < 0.001$) for freshness=0; 0.056 ± 0.005 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: 0.032 ± 0.003 ($p < 0.001$) for overall; 0.002 ± 0.005 ($p = 0.455$) for freshness=0; 0.064 ± 0.005 ($p < 0.001$) for freshness=1. (b) The Pearson correlation of node freshness and originality (disruption) for papers of different team sizes. (e) The Pearson correlation of node freshness and multi-disciplinary impact for papers of different team sizes. For comparison, we calculate the maximum Pearson correlation when we consider team freshness as a weighted linear combination of node and link freshness. We show also the dependence of (c) the mean disruption (originality) D and (f) multi-disciplinarity M on team members' mean career age in 2-author papers. The results suggest that both D and M decrease with team members' mean career age before career age$\leq$ 10 years. The results for the two-tailed Pearson correlation test between D and career age are (coefficients with 95% confidence intervals and p-values): -0.074 ± 0.005 ($p < 0.001$) for overall; -0.030 ± 0.008 ($p < 0.001$) for freshness=0; -0.072 ± 0.007 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: -0.070 ± 0.004 ($p < 0.001$) for overall; -0.002 ± 0.007 ($p = 0.618$) for freshness=0; -0.073 ± 0.006 ($p < 0.001$) for freshness=1.



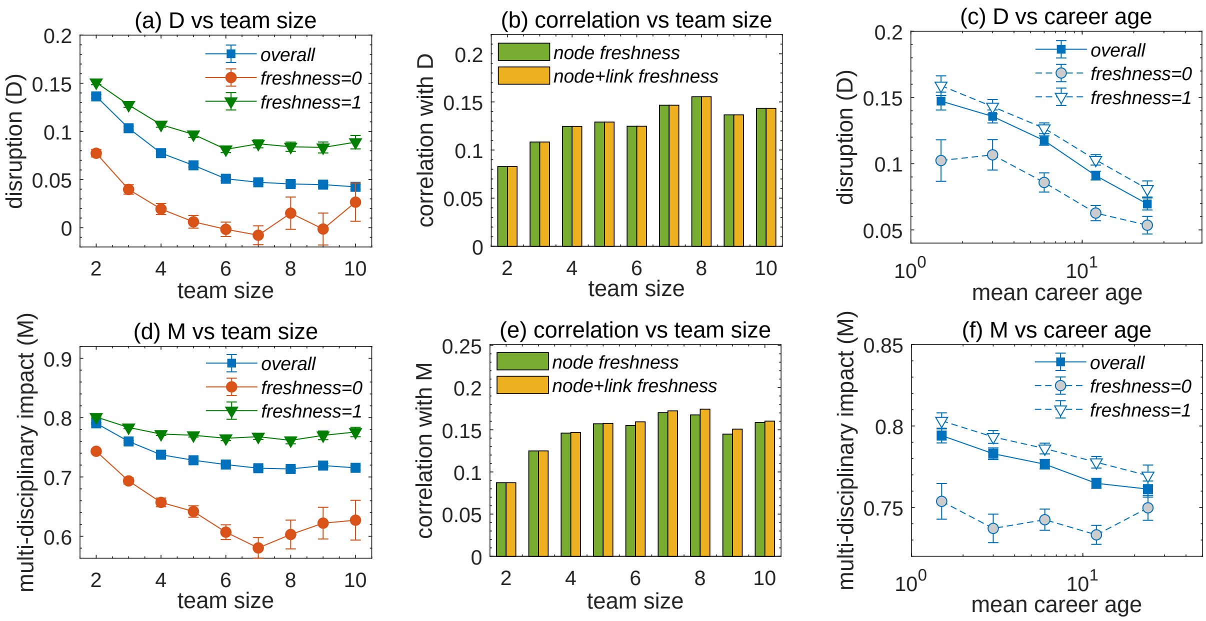
Supplementary Figure 21: **Dependence of disruption (originality) and multi-disciplinarity on team freshness in Chemistry journal (JACS) data.** Shown are the dependence of the disruption (originality) D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. The results suggest that both originality and multi-disciplinarity significantly increase with team freshness. The insets show the distributions of bootstrap disruption or bootstrap multi-disciplinarity. A remarkable difference, i.e., high significance, can be observed between the distributions of D of papers with team freshness 0 and 1. The increasing trend is supported by performing the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.119 ± 0.025 ($p < 0.001$) for 2-author papers; 0.139 ± 0.023 ($p < 0.001$) for 4-author papers; 0.167 ± 0.043 ($p < 0.001$) for 8-author papers. The results for the correlation test between M and team freshness are: 0.114 ± 0.026 ($p < 0.001$) for 2-author papers; 0.138 ± 0.025 ($p < 0.001$) for 4-author papers; 0.128 ± 0.049 ($p < 0.001$) for 8-author papers.



Supplementary Figure 22: **Other results of freshness in Chemistry journal (JACS) data.** Plot of (a) the mean disruption (originality) D and (d) the mean multi-disciplinary impact M of papers versus different team sizes. For each team size, we also study the mean disruption D and the mean multi-disciplinary impact M of papers published by old teams (freshness=0) and fresh teams (freshness=1). The results for the two-tailed Pearson correlation test between D and team size are (coefficients with 95% confidence intervals and p-values): 0.046 ± 0.010 ($p < 0.001$) for overall; -0.006 ± 0.031 ($p = 0.705$) for freshness=0; 0.082 ± 0.015 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: 0.015 ± 0.008 ($p < 0.001$) for overall; -0.035 ± 0.025 ($p = 0.005$) for freshness=0; 0.050 ± 0.013 ($p < 0.001$) for freshness=1. (b) The Pearson correlation of node freshness and originality (disruption) for papers of different team sizes. (e) The Pearson correlation of node freshness and multi-disciplinary impact for papers of different team sizes. For comparison, we calculate the maximum Pearson correlation when we consider team freshness as a weighted linear combination of node and link freshness. We show also the dependence of (c) the mean disruption (originality) D and (f) multi-disciplinarity M on team members' mean career age in 2-author papers. The results for the two-tailed Pearson correlation test between D and career age are (coefficients with 95% confidence intervals and p-values): -0.114 ± 0.025 ($p < 0.001$) for overall; 0.006 ± 0.049 ($p = 0.805$) for freshness=0; -0.102 ± 0.029 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: -0.119 ± 0.026 ($p < 0.001$) for overall; 0.005 ± 0.052 ($p = 0.854$) for freshness=0; -0.118 ± 0.031 ($p < 0.001$) for freshness=1.



Supplementary Figure 23: Dependence of disruption (originality) and multi-disciplinarity on team freshness in the multi-disciplinary data set. This data set contains all papers in five representative multi-disciplinary journals including Nature, Science, Proceedings of the National Academy of Sciences (PNAS), Nature Communications and Science Advances. Shown are the dependence of the disruption (originality) D and multi-disciplinarity M of papers on the team freshness, for (a)(d) 2-author papers, (b)(e) 4-author papers, (c)(f) 8-author papers, respectively. The results suggest that both originality and multi-disciplinarity significantly increase with team freshness. The insets show the distributions of bootstrap disruption or bootstrap multi-disciplinarity. A remarkable difference, i.e., high significance, can be observed between the distributions of D of papers with team freshness 0 and 1. The increasing trend is supported by performing the two-tailed Pearson correlation test between D and team freshness (coefficients with 95% confidence intervals and p-values): 0.083 ± 0.011 ($p < 0.001$) for 2-author papers; 0.125 ± 0.013 ($p < 0.001$) for 4-author papers; 0.155 ± 0.020 ($p < 0.001$) for 8-author papers. The results for the correlation test between M and team freshness are: 0.087 ± 0.011 ($p < 0.001$) for 2-author papers; 0.146 ± 0.013 ($p < 0.001$) for 4-author papers; 0.168 ± 0.022 ($p < 0.001$) for 8-author papers.



Supplementary Figure 24: **Other results of freshness in the multi-disciplinary data set.** This data set contains all papers in five representative multi-disciplinary journals including Nature, Science, Proceedings of the National Academy of Sciences (PNAS), Nature Communications and Science Advances. Plot of (a) the mean disruption (originality) D and (d) the mean multi-disciplinary impact M of papers versus different team sizes. For each team size, we also study the mean disruption D and the mean multi-disciplinary impact M of papers published by old teams (freshness=0) and fresh teams (freshness=1). The results for the two-tailed Pearson correlation test between D and team size are (coefficients with 95% confidence intervals and p-values): -0.113 ± 0.005 ($p < 0.001$) for overall; -0.094 ± 0.017 ($p < 0.001$) for freshness=0; -0.074 ± 0.007 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: -0.092 ± 0.005 ($p < 0.001$) for overall; -0.157 ± 0.017 ($p < 0.001$) for freshness=0; -0.045 ± 0.007 ($p < 0.001$) for freshness=1. (b) The Pearson correlation of node freshness and originality (disruption) for papers of different team sizes. (e) The Pearson correlation of node freshness and multi-disciplinary impact for papers of different team sizes. For comparison, we calculate the maximum Pearson correlation when we consider team freshness as a weighted linear combination of node and link freshness. We show also the dependence of (c) the mean disruption (originality) D and (f) multi-disciplinarity M on team members' mean career age in 2-author papers. The results for the two-tailed Pearson correlation test between D and career age are (coefficients with 95% confidence intervals and p-values): -0.108 ± 0.011 ($p < 0.001$) for overall; -0.051 ± 0.025 ($p < 0.001$) for freshness=0; 0.100 ± 0.012 ($p < 0.001$) for freshness=1. The coefficients between M and team sizes are: -0.075 ± 0.011 ($p < 0.001$) for overall; 0.007 ± 0.025 ($p = 0.565$) for freshness=0; -0.071 ± 0.012 ($p < 0.001$) for freshness=1.

Supplementary Tables

Supplementary Table I: The Pearson correlation coefficient between disruption D (or multi-disciplinary impact M) and team freshness f. We show also the 95% confidence intervals and the p-values of the two-tailed Pearson correlation test. The results suggest that both originality and multi-disciplinarity are positively correlated with team freshness.

Team size	correlation	coefficients	95% confidence intervals	p-values
2-author papers	$c(D, f)$	0.078	$(0.072, 0.084)$	< 0.001
	$c(M, f)$	0.105	$(0.099, 0.111)$	< 0.001
4-author papers	$c(D, f)$	0.105	$(0.096, 0.114)$	< 0.001
	$c(M, f)$	0.146	(0.136,0.156)	< 0.001
8-author papers	$c(D, f)$	0.113	(0.088,0.138)	< 0.001
	$c(M, f)$	0.125	$(0.099, 0.151)$	< 0.001

Supplementary Table II: The Pearson correlation coefficient between disruption D (or multi-disciplinary impact M) and team size s for fresh teams and old teams. We show also the 95% confidence intervals and the p-values of the two-tailed Pearson correlation test. The correlations support the trends observed in Fig. 3 in the main paper.

Correlation	paper set	coefficients	95% confidence intervals	p-values
	overall	-0.027	$(-0.030, -0.023)$	< 0.001
$c(D, s)$	freshness=0	-0.036	$(-0.042, -0.030)$	< 0.001
	freshness=1	0.022	$(0.016, 0.028)$	< 0.001
	overall	-0.005	$(-0.009, -0.001)$	0.013
$c(M, s)$	freshness=0	-0.016	$(-0.022, -0.010)$	< 0.001
	freshness=1	0.055	$(0.062, 0.048)$	< 0.001

Supplementary Table III: The Pearson correlation between link freshness f_l and disruption D (or multi-disciplinary impact M) for each given node freshness f_n in 8-author papers. We show also the 95% confidence intervals and the p-values of the two-tailed Pearson correlation test. The insignificant correlations (large p-values) suggest that for a certain node freshness of a paper, higher link freshness is not associated with higher disruption and multi-disciplinary impact.

Node freshness	correlation	coefficients	95% confidence intervals	p-values
$f_n = 0$	$c(D, f_l)$	-0.004	$(-0.058, 0.050)$	0.898
	$c(M, f_l)$	-0.040	$(0.097, 0.017)$	0.171
$f_n = 0.125$	$c(D, f_l)$	0.000	$(-0.053, 0.053)$	0.996
	$c(M, f_l)$	-0.031	$(-0.088, 0.026)$	0.285
$f_n = 0.25$	$c(D, f_l)$	0.004	$(-0.055, 0.063)$	0.893
	$c(M, f_l)$	0.041	$(-0.022, 0.104)$	0.202
$f_n = 0.375$	$c(D, f_l)$	-0.034	$(-0.101, 0.033)$	0.312
	$c(M, f_l)$	-0.030	$(-0.100, 0.040)$	0.396
$f_n = 0.5$	$c(D, f_l)$	-0.009	$(-0.091, 0.073)$	0.837
	$c(M, f_l)$	0.070	$(-0.017, 0.157)$	0.118
$f_n = 0.625$	$c(D, f_l)$	-0.023	$(-0.123, 0.077)$	0.648
	$c(M, f_l)$	-0.005	$(-0.110, 0.100)$	0.920

Supplementary Table IV: The Pearson correlation coefficient between disruption D (or multi-disciplinary impact M) and mean career age of the team members. We show also the 95% confidence intervals and the p-values of the two-tailed Pearson correlation test. The results suggest that both originality and multi-disciplinarity are negatively correlated with the mean career age of the team members.

Team size	correlation	paper set	coefficients	95% confidence intervals	p-values
	$c(D, age)$	overall	-0.070	$(-0.075, -0.065)$	< 0.001
	$c(D, age)$	freshness=0	-0.046	$(-0.054, -0.038)$	< 0.001
2-author papers	$c(D, age)$	freshness=1	-0.058	$(-0.066, -0.050)$	< 0.001
	$c(M, age)$	overall	-0.090	$(-0.096, -0.084)$	< 0.001
	$c(M, age)$	freshness=0	-0.060	$(-0.069, -0.051)$	< 0.001
	$c(M, age)$	freshness=1	-0.073	$(-0.081, -0.065)$	< 0.001
	$c(D, age)$	overall	-0.077	$(-0.086, -0.068)$	< 0.001
	$c(D, age)$	freshness=0	-0.041	$(-0.057, -0.025)$	< 0.001
4-author papers	$c(D, age)$	freshness=1	-0.054	$(-0.077, -0.031)$	< 0.001
	$c(M, age)$	overall	-0.107	$(-0.116, -0.098)$	< 0.001
	$c(M, age)$	freshness=0	-0.070	$(-0.087, -0.053)$	< 0.001
	$c(M, age)$	freshness=1	-0.077	$(-0.100, -0.054)$	< 0.001
	$c(D, age)$	overall	-0.074	$(-0.099, -0.049)$	< 0.001
	$c(D, age)$	freshness=0	-0.060	$(-0.114, -0.006)$	0.028
8-author papers	$c(D, age)$	freshness=1	-0.013	$(-0.137, 0.111)$	0.844
	$c(M, age)$	overall	-0.069	$(-0.095, -0.043)$	< 0.001
	$c(M, age)$	freshness=0	-0.001	$(-0.057, 0.055)$	0.996
	$c(M, age)$	freshness=1	-0.013	$(-0.144, 0.118)$	0.844

[1] Wu, L., Wang, D. & Evans, J. A. Large teams develop and small teams disrupt science and technology, *Nature* **566**, 378 (2019).

[2] Guimera, R., Uzzi, B., Spiro, J. & Amaral, L. Team assembly mechanisms determine collaboration network structure and team performance, *Science* **308**, 697-702 (2005).

[3] Zhou, T., Lu, L. & Zhang, Y.-C. Predicting missing links via local information, *Eur. Phys. J. B* **71**, 623 (2009).