Supporting Information

Milojević 10.1073/pnas.1309723111

SI Materials and Methods

The following is an alphabetical list of the core journals that were used in the analysis:

Astronomy (four journals): *Astronomical Journal*, *Astronomy & Astrophysics*, *Astrophysical Journal*, and *Monthly Notices of the Royal Astronomical Society.*

Ecology (five journals): *Ecology*, *Journal of Animal Ecology*, *Journal of Ecology*, *Oecologia*, and *Oikos.*

Literature (six journals): *English Literary History*, *English Literature in Transition (1880–1920)*, *Modern Fiction Studies*, *New Literary History*, *Publications of the Modern Language Association*, and *Twentieth Century Literature.*

Mathematics (eight journals): *Acta Mathematica*, *American Journal of Mathematics*, *Annals of Mathematics*, *Inventiones Mathematicae*, *Journal of Functional Analysis*, *Journal of the American Mathematical Society*, *Mathematics of Computation*, and *Proceedings of the London Mathematical Society.*

Social psychology (nine journals): *British Journal of Social Psychology*, *European Journal of Social Psychology*, *Journal of Applied Social Psychology*, *Journal of Experimental Social Psychology*, *Journal of Personality and Social Psychology*, *Journal of Social Psychology*, *Personality and Social Psychology Bulletin*, *Social Behavior and Personality*, and *Social Psychology Quarterly.*

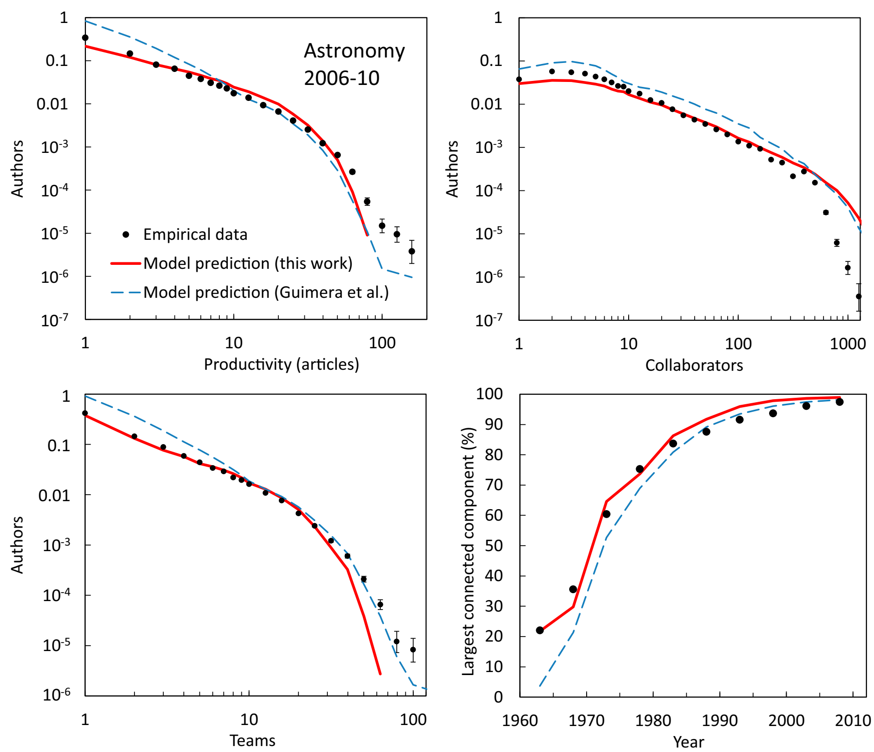

Fig. S1. Comparison of empirical data and model predictions for the distributions of author productivity (*Upper Left*), number of collaborators (*Upper Right*), and teams per author (*Lower Left*), as well as the trend of the relative size of the largest connected component (*Lower Right*), for astronomy (2006–2010). Productivity (number of articles per author) is determined based on all articles regardless of author role (lead author or coauthor). Collaborators are defined as all coauthors of a given author from all articles published during this period. The number of teams with which a given author is involved is determined as the number of different lead authors in all of the papers in which the given author is a coauthor. The largest connected component is the fraction of all authors that are linked by coauthorship in this time period. Values above 50% signal the emergence of a giant component. Model predictions are given for the model featured in this work (red solid lines) and, as a comparison, based on the principles of team assembly laid out in ref. 1 (blue dashed lines). The model by Guimerà et al. requires explicit input on team size for each article, which we take from our model. Then, for the Guimerà et al. predictions, we choose team members by assuming probability p that the team member is an author already present in the network (incumbent), and probability q that the incumbent is already a collaborator of some other team member. We use $p = 0.78$ and $q = 0.82$, as appropriate for astronomy (ref. 1, figure 4*J*). Also, to recreate the Guimerà et al. process fully, we implement removal of authors who have been inactive for longer than 5,000 time steps (ref. 1, supplemental information). Both the current and Guimerà et al. methods for team buildup provide good general description of author-centric distributions and of the evolution of coauthorship network topology.

1. Guimerà R, Uzzi B, Spiro J, Amaral LA (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308(5722):697–702.

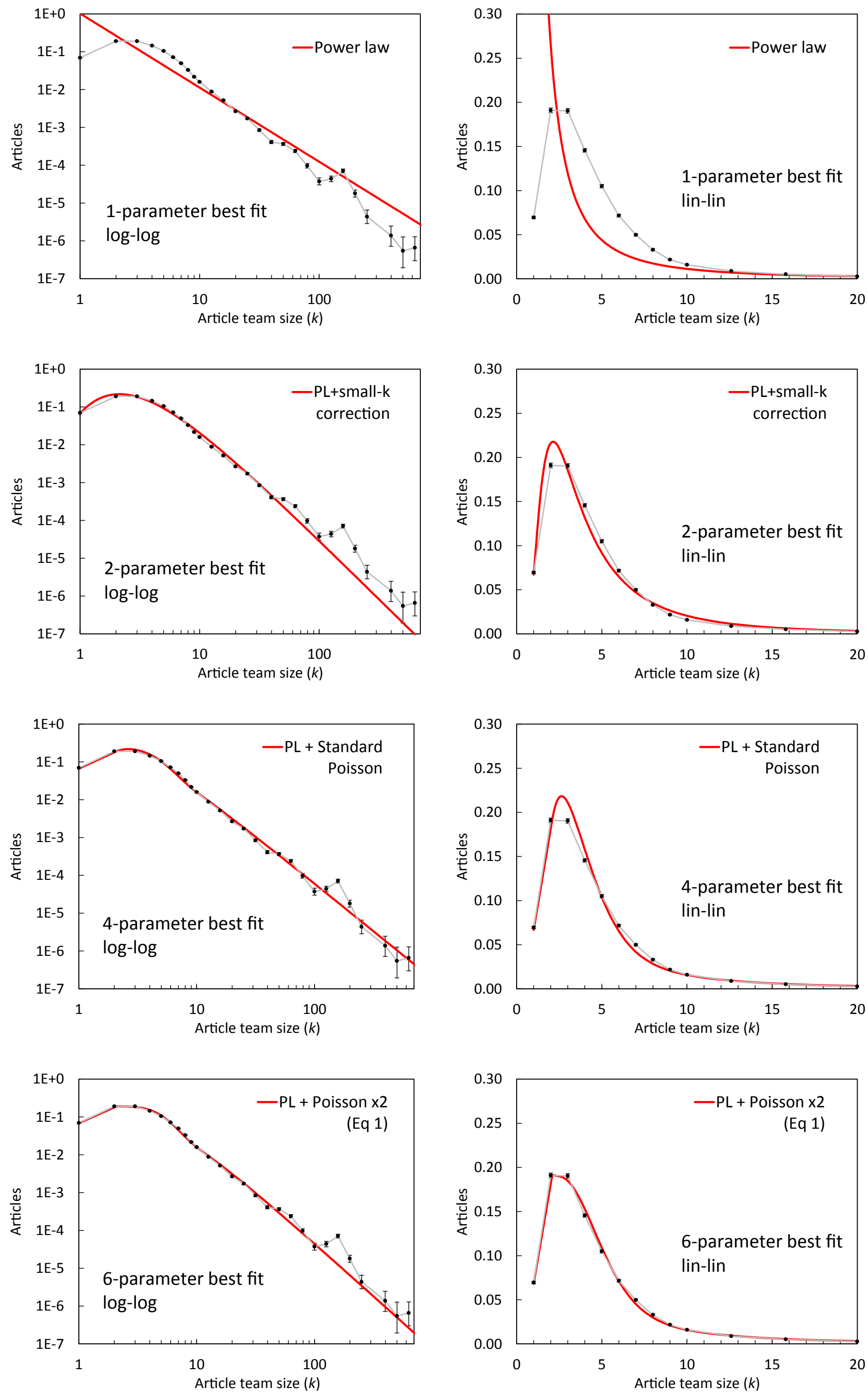

Fig. S2. Number of parameters used for the functional description of team size distribution and the resulting quality of fits (astronomy, 2006–2010). The left-hand panels show distributions on a log-log scale, whereas the right-hand panels show the small-*k* regions on a linear-linear scale. The first row shows the best fit if one assumes only the power law, i.e., a single-parameter fit (plus a normalization), which is clearly inadequate at $k < 10$. The second row shows the best fit when the power law is modified with a small-*k* correction [in the form $\exp(-\beta/k)\ k^{-\alpha}$], which requires two parameters. The fit is improved but fails to acknowledge that the small-*k* "hook" sits partially above the extrapolation of the power-law tail. Furthermore, this modification of the power law cannot fit the distributions at earlier time periods that are dominated by the Poisson distribution. The third row adds a Poisson component, requiring a total of four parameters. The fit is now significantly improved, but there are still some discrepancies at small *k* due to the recent rise of a different flavor of Poisson teams that always have at least two authors (presumably built around the student–mentor pair; see main article for details). Modeling this deviation requires another Poisson function, which adds two parameters (fourth row) and becomes Eq. **1**. This additional Poisson component is very prominent in the distribution of team sizes of first-time authors, and in the overall distribution of some disciplines (e.g., mathematics and social psychology; Table 1).

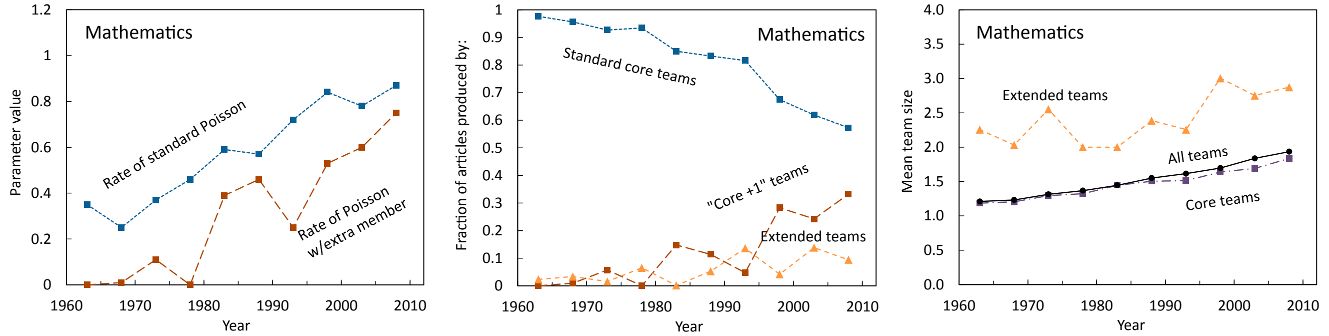

Fig. S3. Trends in team evolution in mathematics (1961–2010). Refer to the legend of Fig. 6 for details. The slope of the power-law component is poorly constrained and was omitted from the left panel.

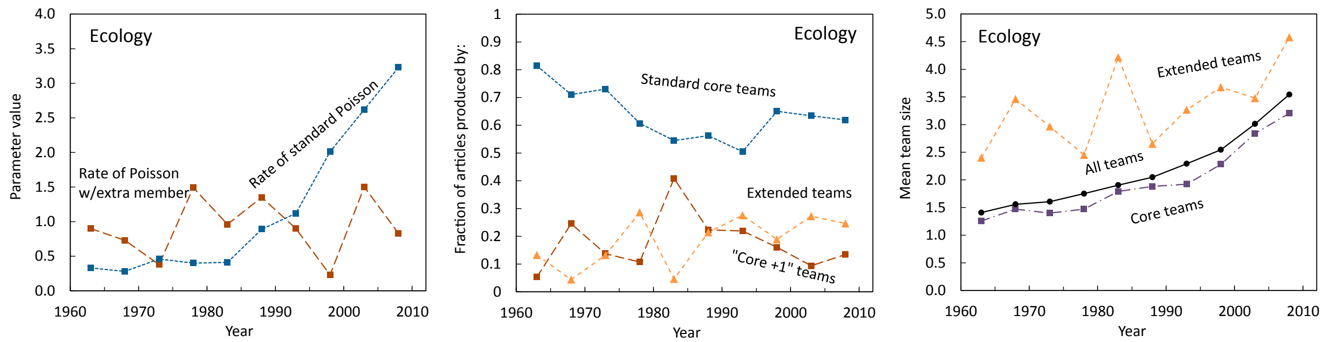

Fig. S4. Trends in team evolution in ecology (1961–2010). Refer to legend of Fig. 6 for details. The slope of the power-law component is poorly constrained and was omitted from the left panel.