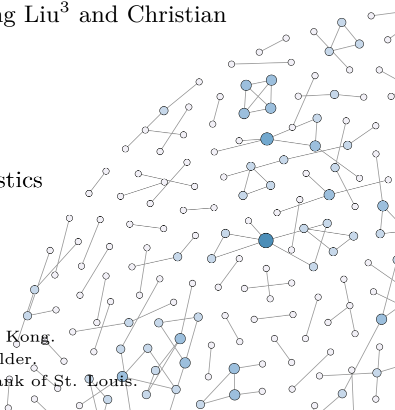
Superstar Economists: Coauthorship networks and research output

Michael D. König[1]
joint with Chih-Sheng Hsie[2], Xiaodong Liu[3] and Christian Zimmermann[4]

LMU Munich
Department of Statistics

18th April 2018



[1] Department of Economics, University of Zurich.
[2] Department of Economics, Chinese University of Hong Kong.
[3] Department of Economics, University of Colorado Boulder.
[4] Department of Economic Research, Federal Reserve Bank of St. Louis.

Contribution: Theory

- We build a micro-founded model for the output produced in scientific co-authorship networks.[5]
- Differently to previous works, we are able to characterize the interior equilibrium when multiple agents spend effort in multiple, possibly overlapping projects, and there are interaction effects in the cost of effort.
- The equilibrium solution then allows us to study the impact of individual researchers on total research output (ranking), and the optimal design of research funding programs.

[5] Coralio Ballester, Antoni Calvó-Armengol, and Yves Zenou. "Who's Who in Networks. Wanted: The Key Player". *Econometrica* 74.5 (2006), pp. 1403–1417; Antonio Cabrales, Antoni Calvó-Armengol, and Yves Zenou. "Social interactions and spillovers". *Games and Economic Behavior* 72.2 (2011), pp. 339–360; Matthew O. Jackson and Asher Wolinsky. "A Strategic Model of Social and Economic Networks". *Journal of Economic Theory* 71.1 (1996), pp. 44–74.

Contribution: Empirics

- We develop a novel estimation framework in which agents can contribute to many potentially overlapping projects, and the participation is endogenously modelled.
- The allocation of agents into different projects is determined by a matching process that depends on both, the authors' and projects' characteristics,[6] while the effort levels are determined in the Nash equilibrium.
- We estimate this model using data for the network of scientific coauthorships between economists registered in the Research Papers in Economics (RePEc) author service.[7]

[6] Arun Chandrasekhar. "Econometrics of network formation". In: *The Oxford Handbook of the Economics of Networks.* 2015, pp. 303–357; A. Chandrasekhar and M. Jackson. "Tractable and Consistent Random Graph Models". *Available at SSRN 2150428* (2012).

[7] http://repec.org/

Contribution: Policy

- We develop a novel ranking measure (for economists and their departments) that quantifies the endogenous decline in research output due to the removal of an economist from the network ("key players", "superstar" economists).[8]
- We find that the highest ranked authors are not necessarily the ones with the largest number of citations, or coincide with other ranking measures used in the literature.
- However, this discrepancy is not surprising, as traditional rankings are typically not derived from microeconomic foundations, and do not take into account the spillover effects generated in scientific knowledge production networks.

---

[8] Yves Zenou. "Key Players". *Oxford Handbook on the Economics of Networks, Y. Bramoulle, B. Rogers and A. Galeotti (Eds.), Oxford University Press* (2015); Fabian Waldinger. "Peer effects in science: evidence from the dismissal of scientists in Nazi Germany". *The Review of Economic Studies* 79.2 (2012), pp. 838–861; Pierre Azoulay, Joshua Graff Zivin, and Jialan Wang. "Superstar Extinction". *The Quarterly Journal of Economics* 125.2 (2010), pp. 549–589.

- Our model further allows us to solve an optimal research funding problem of a planner who wants to maximize total scientific output by introducing research grants into the author's payoff function.[9]
- A comparison of our optimal funding policy with the National Science Foundation (NSF) indicates significant differences, both at the individual and the departmental levels (e.g. NSF awards are uncorrelated with the number of coauthors).
- Our results highlight the importance of the coauthorship network in determining the optimal funding policy, while it does not seem to have an effect on the research funding program by the NSF.

[9] Paula E Stephan. "The economics of science". *Journal of Economic literature* (1996), pp. 1199–1235; Paula E. Stephan. *How economics shapes science*. Harvard University Press, 2012.

Production Function

- Assume that there are $s \in \mathcal{P} = \{1, \ldots, p\}$ research projects (papers) and $i \in \mathcal{N} = \{1, \ldots, n\}$ researchers (authors).
- Let the *production function* for project s be given by

$$Y_s = Y_s(\mathcal{G}) = \sum_{i \in \mathcal{N}} \alpha_i e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} f_{ij} e_{is} e_{js}, \tag{1}$$

- where Y_s is the research output of project s, e_{is} is the research effort that agent i spent in project s ($e_{is} = 0$ if agent i does not participate in project s),
- α_i captures the productivity of agent i,
- $f_{ij} \in (0, 1]$ measures the (knowledge) similarity between agents i and j,
- the spillover-effect parameter $\lambda > 0$ represents complementarity between the research efforts of collaborating agents, and
- $\mathcal{G}$ represents the *bipartite* network of authors and projects.

Example

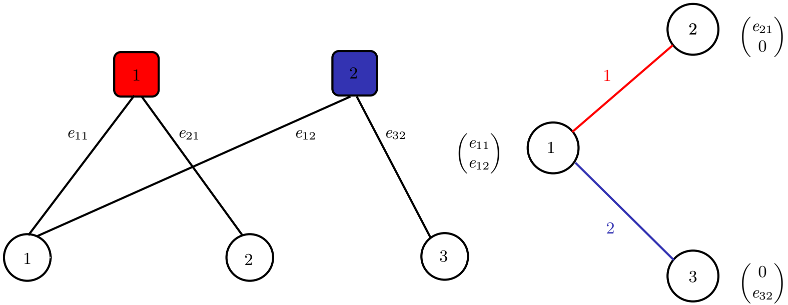


Figure: (Left panel) The bipartite collaboration network $\mathcal{G}$ of authors and projects, where round circles represent authors and squares represent projects. (Right panel) The projection of the bipartite network $\mathcal{G}$ on the set of coauthors.

Utility

- The utility of agent i is then given by

$$U_i = U_i(\mathcal{G}) = \underbrace{\sum_{s \in \mathcal{P}} g_{is} \delta_s Y_s}_{\text{payoff}} - \underbrace{\frac{1}{2} \left(\sum_{s \in \mathcal{P}} e_{is}^2 + \phi \sum_{s \in \mathcal{P}} \sum_{t \in \mathcal{P} \backslash \{s\}} e_{is} e_{it} \right)}_{\text{cost}}, \qquad (2)$$

- where $g_{is} \in \{0, 1\}$ indicates whether agent i participates in project s, $\delta_s \in (0, 1]$ is a discount factor,[10] and
- the parameter $\phi > 0$ represents substitutability between the research efforts of the same agent in different projects.

[10] If $\delta_s = 1$, then individual payoff from research output Y_s is not discounted. If $\delta_s = 1/\sum_{i \in \mathcal{N}} g_{is}$, then individual payoff is discounted by the number of agents (coauthors) participating in project s.

Definitions

- Let

$$\mathbf{W} = \mathbf{G}(\mathrm{diag}_{s=1}^{p}\{\delta_s\} \otimes \mathbf{F})\mathbf{G}, \qquad \text{and} \qquad \mathbf{M} = \mathbf{G}(\mathbf{J}_p \otimes \mathbf{I}_n)\mathbf{G}, \tag{3}$$

  where $\otimes$ denotes Kronecker product, $\mathbf{G}$ is an np-dimentional diagonal matrix given by $\mathbf{G} = \mathrm{diag}_{s=1}^{p}\{\mathrm{diag}_{i=1}^{n}\{g_{is}\}\}$,
- $\mathbf{F}$ is an $n \times n$ zero-diagonal matrix with the (i, j)-th $(i \neq j)$ element being f_{ij}, and
- $\mathbf{J}_p$ is an $p \times p$ zero-diagonal matrix with off-diagonal elements being ones.
- Further, let $\rho_{\max}(\mathbf{A})$ denote the spectral radius of a square matrix $\mathbf{A}$.

Equilibrium Characterization

- **Proposition:** Suppose the production function for each project $s \in \mathcal{P}$ is given by Equation (1) and the utility function for each agent $i \in \mathcal{N}$ is given by Equation (2). Given the bipartite network $\mathcal{G}$, if

$$|\lambda| < 1/\rho_{\max}(\mathbf{W}) \qquad \text{and} \qquad |\phi| < 1/\rho_{\max}((\mathbf{I}_{np} - \lambda \mathbf{W})^{-1}\mathbf{M}), \tag{4}$$

then the equilibrium effort portfolio is given by

$$\mathbf{e}^* = (\mathbf{I}_{np} - \mathbf{L}^{\lambda,\phi})^{-1}\mathbf{G}(\boldsymbol{\delta} \otimes \boldsymbol{\alpha}), \tag{5}$$

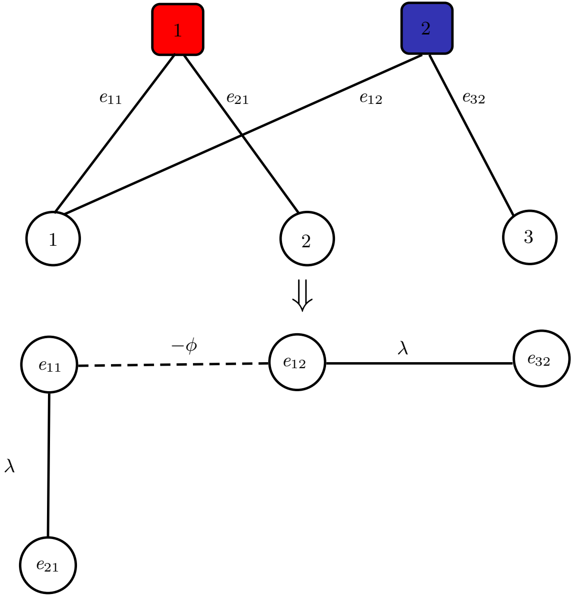
where $\mathbf{L}^{\lambda,\phi} = \lambda \mathbf{W} - \phi \mathbf{M}$, $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_p]'$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_n]'$.

Line Graph

- **Definition:** Given a graph $\mathcal{G}$, its *line graph* $L(\mathcal{G})$ is a graph such that each node of $L(\mathcal{G})$ represents an edge of $\mathcal{G}$, and two nodes of $L(\mathcal{G})$ are adjacent if and only if their corresponding edges share a common endpoint in $\mathcal{G}$.[11]
- Then the matrix $\mathbf{L}^{\lambda,\phi}$ represents a weighted adjacency matrix of the *line graph* $L(\mathcal{G})$ for the bipartite network $\mathcal{G}$, where
  - each link between nodes sharing a project has weight $\lambda\delta_s f_{ij}$, and
  - each link between nodes sharing an author has weight $-\phi$.

[11] Cf. e.g. Douglas B. West. *Introduction to Graph Theory*. 2nd. Prentice-Hall, 2001.

Example


1
2
e_{11}
e_{21}
e_{12}
e_{32}
1
2
3
$\Downarrow$
e_{11}
$-\phi$
e_{12}
λ
e_{32}
λ
e_{21}

- Following Equation (3),

$$\mathbf{W} = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{M} = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and, hence,

$$\mathbf{L}^{\lambda,\phi} = \lambda\mathbf{W} - \phi\mathbf{M} = \begin{bmatrix} 0 & \lambda & 0 & -\phi & 0 & 0 \\ \lambda & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ -\phi & 0 & 0 & 0 & 0 & \lambda \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & \lambda & 0 & 0 \end{bmatrix}.$$

- The sufficient condition for the existence of a unique equilibrium given by (4) holds if $|\lambda| < 1$ and $|\phi| < 1 - \lambda^2$.
- From Equation (5) the equilibrium effort portfolio is

$$\mathbf{e}^* = \begin{bmatrix} e_{11}^* \\ e_{21}^* \\ e_{31}^* \\ e_{12}^* \\ e_{22}^* \\ e_{32}^* \end{bmatrix} = (\mathbf{I}_{np} - \mathbf{L}^{\lambda,\phi})^{-1}\mathbf{G}(\mathbf{I}_{np} \otimes \boldsymbol{\alpha})$$

$$= \frac{1}{(1-\lambda^2)^2 - \phi^2} \begin{bmatrix} (1-\lambda^2-\phi)\alpha_1 + \lambda(1-\lambda^2)\alpha_2 - \lambda\phi\alpha_3 \\ \lambda(1-\lambda^2-\phi)\alpha_1 + (1-\lambda^2-\phi^2)\alpha_2 - \lambda^2\phi\alpha_3 \\ 0 \\ (1-\lambda^2-\phi)\alpha_1 - \lambda\phi\alpha_2 + \lambda(1-\lambda^2)\alpha_3 \\ 0 \\ \lambda(1-\lambda^2-\phi)\alpha_1 - \lambda^2\phi\alpha_2 + (1-\lambda^2-\phi^2)\alpha_3 \end{bmatrix}.$$

- Observe that

$$
\begin{aligned}
\frac{\partial e_{11}^*}{\partial \alpha_1} &= \frac{\partial e_{12}^*}{\partial \alpha_1} = \frac{1}{1-\lambda^2+\phi} > 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_1} &= \frac{\partial e_{32}^*}{\partial \alpha_1} = \frac{\lambda}{1-\lambda^2+\phi} > 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_2} &= \frac{\partial e_{32}^*}{\partial \alpha_3} = \frac{1-\lambda^2-\phi^2}{(1-\lambda^2)^2-\phi^2} > 0 \\
\frac{\partial e_{11}^*}{\partial \alpha_2} &= \frac{\partial e_{12}^*}{\partial \alpha_3} = \frac{\lambda(1-\lambda^2)}{(1-\lambda^2)^2-\phi^2} > 0
\end{aligned}
$$

  which suggest that more productive agents raise not only their own effort levels but also the effort levels of their collaboration partners.

- On the other hand,

$$
\begin{aligned}
\frac{\partial e_{11}^*}{\partial \alpha_3} &= \frac{\partial e_{12}^*}{\partial \alpha_2} = -\frac{\lambda\phi}{(1-\lambda^2)^2-\phi^2} < 0 \\
\frac{\partial e_{21}^*}{\partial \alpha_3} &= \frac{\partial e_{32}^*}{\partial \alpha_2} = -\frac{\lambda^2\phi}{(1-\lambda^2)^2-\phi^2} < 0
\end{aligned}
$$

  which suggest that more productive agents induce lower effort levels spent by agents on other projects.

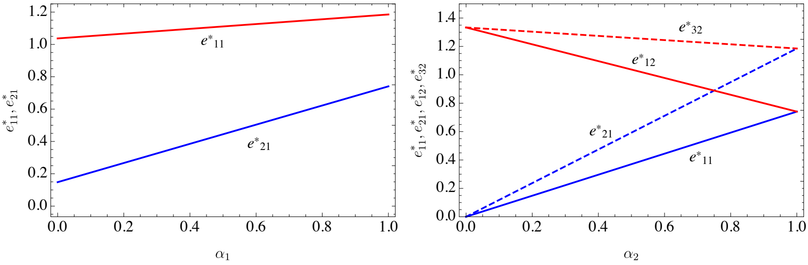


FIGURE: (Left panel) Equilibrium effort levels for agents 1 and 2 in project 1 for $\phi = 0.75$, $\lambda = 0.25$, $\alpha_2 = \alpha_3 = 1$ (where $e^*_{11} = e^*_{12}$ and $e^*_{21} = e^*_{32}$) and varying values of α_1. (Right panel) Equilibrium effort levels for agents 1, 2 and 3 in projects 1 and 2 for $\alpha_1 = \alpha_3 = 1$, $\phi = 0.75$, $\lambda = 0.25$ and varying values of α_2.

- The marginal change of the equilibrium effort e_{11}^* of agent 1 in project 1 with respect to the spillover parameter λ is given by

$$\frac{\partial e_{11}^*}{\partial \lambda} = \frac{1}{\left((1-\lambda^2)^2-\phi^2\right)^2}\Big(2\lambda(1-\lambda^2-\phi)^2\alpha_1$$
$$+((1-\lambda^4-\phi^2)(1-\lambda^2)+2\lambda^2\phi^2)\alpha_2 - \phi((1+3\lambda^2)(1-\lambda^2)-\phi^2)\alpha_3\Big).$$

- Observe that the coefficient of α_3 is negative. Thus, when the ability α_3 is large enough, $\partial e_{11}^*/\partial \lambda$ could be negative.
- Similarly, when the substitution effect ϕ is large, agent 1 may spend even less effort in the project with agent 2, indicating congestion effects across projects.

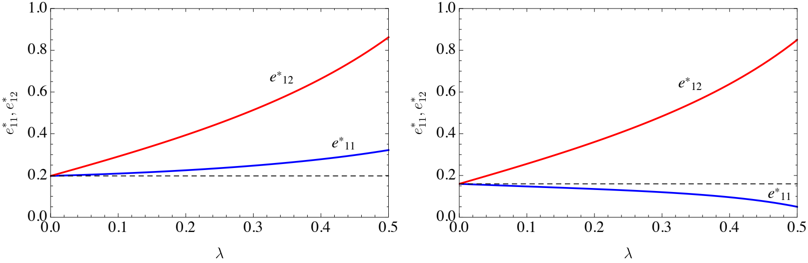


FIGURE: Equilibrium effort levels for agent 1 with $\alpha_1 = 0.2$, $\alpha_2 = 0.1$, $\alpha_3 = 0.9$, $\phi = 0.05$ (left panel) and $\phi = 0.25$ (right panel) for varying values of λ. The dashed lines in the bottom panels indicate the effort level for $\lambda = 0$.

Superstars, Key Players and Rankings

- We analyze the impact of the removal of individual authors from the coauthorship network on overall scientific output.[12]
- The *key author* is defined by

$$i^* \equiv \operatorname*{argmax}_{i \in \mathcal{N}} \left\{ \sum_{s \in \mathcal{P}} Y_s(\mathcal{G}) - \sum_{s \in \mathcal{P}} Y_s(\mathcal{G} \backslash \{i\}) \right\}. \tag{6}$$

- Further, aggregating researchers to their departments $\mathcal{D} \subset \mathcal{N}$ allows us to compute the *key department* as

$$\mathcal{D}^* \equiv \operatorname*{argmax}_{\mathcal{D} \subset \mathcal{N}} \left\{ \sum_{s \in \mathcal{P}} Y_s(\mathcal{G}) - \sum_{s \in \mathcal{P}} Y_s(\mathcal{G} \backslash \mathcal{D}) \right\}. \tag{7}$$

[12] Fabian Waldinger. "Quality matters: The expulsion of professors and the consequences for PhD student outcomes in Nazi Germany". *Journal of Political Economy* 118.4 (2010), pp. 787–831.

Research Funding

- We consider a two-stage game:
  - In the first stage, the planner announces the research funding scheme $r \in \mathbb{R}^n_+$ that the authors should receive, and
  - in the second stage the authors choose their research efforts, given r.
- The optimal funding profile r^* can then be found by backward induction.
- Consider the second stage. We assume that agent $i \in \mathcal{N}$ receives research funding, $r \geq 0$, proportional to the output she generates:

$$U_i(\mathcal{G}, r) = \sum_{s \in \mathcal{P}} g_{is}\delta_s Y_s - \frac{1}{2}\left(\sum_{s \in \mathcal{P}} e_{is}^2 + \phi \sum_{s \in \mathcal{P}} \sum_{t \in \mathcal{P} \setminus \{s\}} e_{is}e_{it}\right) + \underbrace{r \sum_{s \in \mathcal{P}} g_{is}\delta_s Y_s}_{\text{research funding}} \tag{8}$$

- **Proposition:** Suppose the production function for each project $s \in \mathcal{P}$ is given by Equation (1) and the utility function for each agent $i \in \mathcal{N}$ is given by Equation (8). Given the bipartite network $\mathcal{G}$, if

$$|\lambda| < 1/((1+r)\rho_{\max}(\mathbf{W})) \qquad \text{and} \qquad |\phi| < 1/\rho_{\max}((\mathbf{I}_{np} - (1+r)\lambda\mathbf{W})^{-1}\mathbf{M}), \tag{9}$$

then the equilibrium effort portfolio is given by

$$\mathbf{e}^*(r) = (\mathbf{I}_{np} - \mathbf{L}_r^{\lambda,\phi})^{-1}\mathbf{G}(\boldsymbol{\delta} \otimes \boldsymbol{\alpha}), \tag{10}$$

where $\mathbf{L}_r^{\lambda,\phi} = \lambda(1+r)\mathbf{W} - \phi\mathbf{M}$, $\delta = [\delta_1, \cdots, \delta_p]'$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_n]'$.

Planner's Problem

- Given the equilibrium effort portfolio, in the first stage of the game, the planner maximizes total output, $\sum_{s\in\mathcal{P}} Y_s$, less total cost of the policy, $r\sum_{s\in\mathcal{P}}\sum_{i\in\mathcal{N}} g_{is}\delta_s Y_s$.
- The planner's problem can thus be written as

$$r^* = \underset{r\in\mathbb{R}_+}{\operatorname{argmax}} \sum_{s\in\mathcal{P}} \left(Y_s(\mathcal{G}, r) - r\sum_{i\in\mathcal{N}} g_{is}\delta_s Y_s(\mathcal{G}, r) \right), \tag{11}$$

- where $Y_s(\mathcal{G}, r)$ is the output of project s from Equation (1) with the equilibrium effort levels $\mathbf{e}^*(r)$ given by Equation (10).
- Equation (11) can then be solved numerically using a fixed point algorithm.

Data

- The data used for this study makes extensive use of the metadata assembled by the *Research Papers in Economics (RePEc)* author service initiative and its various projects.[13]
- RePEc assembles the information about publications relevant to economics from 1900 publishers, including all major commercial publishers and university presses, policy institutions and pre-prints from academic institutions.
- At the time of this writing, this encompasses 2.2 million records, including 0.75 million for pre-prints.
- We take the publication profiles of economists registered with the RePEc Author Service (49,000 authors), that includes what they have published and where they are affiliated.[14]

[13] http://repec.org/

[14] https://authors.repec.org/

- We get information about their *advisors*, students and alma mater, as recorded in the RePEc Genealogy project.[15]
- We gather in which mailing lists the papers have been disseminated through the NEP project.[16] The latter have human editors determining to which *field* new working papers belong.
- We make use of paper *download data* that is made available by the LogEc project.[17]
- We use *citations* to the papers and articles as extracted by the CitEc project.[18]
- We use journal *impact factors*, author and institution rankings from IDEAS.[19]
- We make use of the "Ethnea" tool at the University of Illinois to establish the *ethnicity* of authors based on the first and last names.[20]

data

[15] https://genealogy.repec.org/

[16] https://nep.repec.org/

[17] http://logec.repec.org/

[18] http://citec.repec.org/

[19] https://ideas.repec.org/top/. For a detailed description of the factors and rankings, see Zimmermann (2013).

[20] http://abel.lis.illinois.edu/cgi-bin/ethnea/search.py

Summary Statistics

Table: Summary statistics for the 2010-2012 selected sample.

	Min	Max	Mean	S.D.	Sample size
Papers					
Citation recursive Impact Factor	0.0000	115.5851	6.5796	12.2021	3620
number of authors (in each paper)	1	5	1.8892	0.7108	3620
Authors					
Log life-time citations	0	10.5516	5.4948	1.7118	1925
Decades after Ph.D. graduation	-0.6	5.9000	1.1113	0.9909	1925
Female	0	1	0.1345	0.3413	1925
NBER connection	0	1	0.1195	0.3244	1925
Ivy League connection	0	1	0.1553	0.3623	1925
Editor	0	1	0.0494	0.2167	1925
number of papers (for each author)	1	74	3.5527	3.8339	1925

Note: We drop authors who did not coauthor with any others during the sample period. We also drop papers without any citations when extracting from the RePEc data base.

Distributions of Authors, Papers and Quality

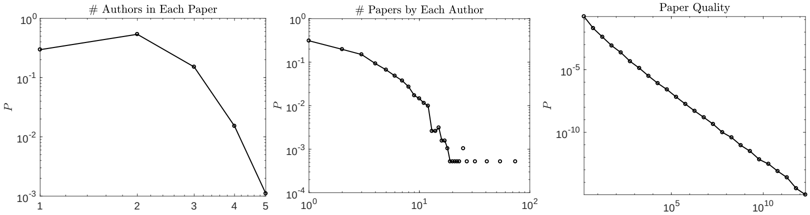
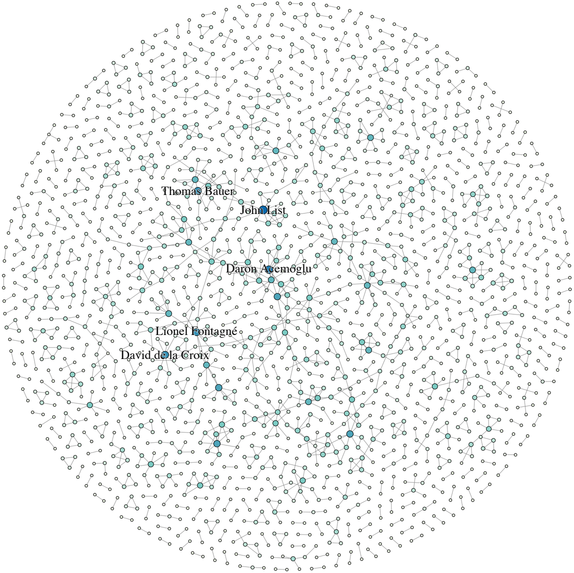
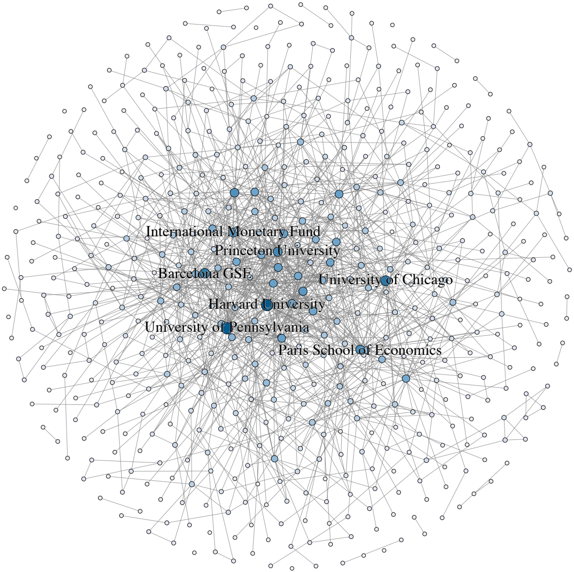


Figure: The distribution of authors per paper (left panel), the number of papers per author (middle panel) and the paper quality (right panel).

Collaboration Network of Authors



Collaboration Network of Departments



Estimating the Production Function

- Suppose there are n authors and p papers. Following Equation (1), the production function of paper s, with $s = 1, \ldots, p$, is given by

$$Y_s = \sum_{i \in \mathcal{N}} \alpha_i e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} f_{ij} e_{is} e_{js} + \epsilon_s, \tag{12}$$

- where ϵ_s is a paper-specific random shock.
- We assume $\alpha_i = \mathbf{x}_i' \beta$, where $\mathbf{x}_i$ is a $k \times 1$ vector of author-specific exogenous characteristics.
- The empirical production function can be estimated by the the nonlinear least squares (NLS) method or the maximum likelihood (ML) method (under the normality assumption on ϵ_s), with the unobservable e_{is} replaced by the equilibrium research effort given in Equation (5).

Matching Process

- A problem with directly estimating Equation (12) is the potential *endogeneity* of $\mathbf{G} = \text{diag}_{s=1}^{p}\{\text{diag}_{i=1}^{n}\{g_{is}\}\}$.
- To address this endogeneity problem, we model the endogenous matching process of author i to paper s by

$$g_{is} = \mathbb{1}_{\{\psi_{is}+u_{is}>0\}}, \tag{13}$$

  with

$$\psi_{is} = \mathbf{z}_{is}'\gamma_1 + \gamma_2\mu_i + \gamma_3\kappa_s,$$

  and $\mathbf{z}_{is}$ denoting an $h \times 1$ vector of dyad-specific regressors, capturing the similarity between author i and paper s, measured by the research overlap in the NEP fields of paper s and author i.
- The identification of the spillover parameter λ then comes from the exogenous variation in the research overlap between author i and project s (cf. Ductor, 2014).
- The variable μ_i represents author i's unobserved characteristic; and κ_s represents the paper's unobserved characteristic (Graham, 2016).

- The production function (12) can then be extended to

$$Y_s = \sum_{i \in \mathcal{N}} (\underbrace{\mathbf{x}_i' \beta + \zeta \mu_i}_{\alpha_i}) e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} f_{ij} e_{is} e_{js} + \underbrace{\eta \kappa_s + v_s}_{\epsilon_s}, \qquad (14)$$

  where v_s is assumed to be independent of u_{is} and normally distributed with zero mean and variance σ_v^2.

- Given $\mathbf{X} = [\mathbf{x}_i]$ and $\mathbf{Z} = [\mathbf{z}_{is}]$, the joint probability function of $\mathbf{Y} = (Y_1, \cdots, Y_p)$ and $\mathbf{G}$ can be specified as

$$\Pr(\mathbf{Y}, \mathbf{G} | \mathbf{X}, \mathbf{Z}) = \int_\mu \int_\kappa \Pr(\mathbf{Y} | \mathbf{G}, \mathbf{X}, \mathbf{Z}, \mu, \kappa) \Pr(\mathbf{G} | \mathbf{Z}, \mu, \kappa) f(\mu) f(\kappa) \, d\mu \, d\kappa, \qquad (15)$$

  from which we can estimate the parameter vector $\theta = (\lambda, \phi, \beta', \gamma', \eta, \zeta, \sigma_v^2)'$, with $\gamma = (\gamma_1', \gamma_2, \gamma_3)'$.

Estimation Results

Table: Estimation results for the 2010-2012 sample.[a]

		Homogeneous Spillovers		Heterogeneous Spillovers		Discounting # Coauthors	
		Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Matching							
Constant	(γ_0)	–	-8.3796***	–	-8.2791***	–	-8.3396***
			(0.0397)		(0.0413)		(0.0421)
Same NEP	(γ_1)	–	5.9872***	–	5.8110***	–	5.8971***
			(0.0715)		(0.0706)		(0.0735)
Author effect	(γ_2)	–	2.2147***	–	1.8968***	–	2.0662***
			(0.0649)		(0.0658)		(0.0620)
Project effect	(γ_3)	–	-0.6211***	–	-0.0031	–	-0.0216***
			(0.1178)		(0.1117)		(0.1244)
Sample size (papers)		3,620		3,620		3,620	
Sample size (authors)		1,925		1,925		1,925	

[a] Model (1): Assuming exogenous matching between authors and papers. Model (2): Assuming endogenous matching by Equation (13). The asterisks ***(**,*) indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.

TABLE: Estimation results for the 2010-2012 sample.[a]

		Homogeneous Spillovers		Heterogeneous Spillovers		Discounting # Coauthors	
		Model (1)	Model (2)	Model (1)	Model (2)	Model (1)	Model (2)
Output							
Spillover	(λ)	-0.1084***	0.0239**	-0.0954**	0.1305***	-0.1883***	0.1259***
		(0.0340)	(0.0117)	(0.0479)	(0.0295)	(0.0594)	(0.0447)
Cost	(ϕ)	0.0068	0.8192***	0.0057	0.4928***	0.0070	0.7949***
		(0.0054)	(0.0684)	(0.0048)	(0.0523)	(0.0051)	(0.0797)
Constant	(β_0)	-0.6712***	-1.3754***	-0.7263***	-1.4636***	-0.6508***	-1.5150***
		(0.1322)	(0.1141)	(0.1402)	(0.1344)	(0.1304)	(0.1312)
Log life-time citat.	(β_1)	0.2475***	0.3203***	0.2477***	0.3395***	0.2450***	0.3457***
		(0.0227)	(0.0171)	(0.0238)	(0.0209)	(0.0222)	(0.0178)
Decades after grad.	(β_2)	-0.2497***	-0.3673***	-0.2417***	-0.4070***	-0.2463***	-0.4225***
		(0.0400)	(0.0310)	(0.0414)	(0.0317)	(0.0401)	(0.0256)
Female	(β_3)	0.1871***	-0.0984*	0.1892***	0.1867***	0.1857***	0.3119***
		(0.0656)	(0.0579)	(0.0681)	(0.0541)	(0.0647)	(0.0542)
NBER connection	(β_4)	0.2779***	0.6629***	0.2784***	0.5185***	0.2773***	0.5604***
		(0.0506)	(0.0321)	(0.0518)	(0.0298)	(0.0520)	(0.0354)
Ivy League connect.	(β_5)	0.1788***	0.3107***	0.1885***	0.1854***	0.1824***	0.2359***
		(0.0455)	(0.0347)	(0.0472)	(0.0371)	(0.0465)	(0.0461)
Editor	(β_6)	-0.1107	-1.3971***	-0.1010	-0.7778***	-0.1088	-0.9342***
		(0.0919)	(0.0971)	(0.0950)	(0.0885)	(0.0915)	(0.1454)
Author effect	(ζ)	–	2.7654***	–	2.2938***	–	2.5636***
			(0.0884)		(0.0819)		(0.0926)
Project effect	(η)	–	1.2202**	–	0.2521	–	0.0189
			(0.6687)		(0.6164)		(0.6728)
Project variance	(σ_v^2)	118.5337***	67.8102***	119.0462***	70.1867***	118.6615***	68.1037***
		(2.8501)	(1.6603)	(2.8229)	(1.7446)	(2.8154)	(1.7650)
Sample size (papers)		3,620		3,620		3,620	
Sample size (authors)		1,925		1,925		1,925	

[a] Model (1): Assuming exogenous matching between authors and papers. Model (2): Assuming endogenous matching by Equation (13). The asterisks ***(**,*) indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.

Robustness: Assortative matching

- We further enrich the matching equation to capture assortative matching among researchers.[21]
- We use the average characteristics of the authors in each project as the additional characterizations of the project.
- We then can extend the control variable $\mathbf{z}_{is}$ between project s and each author i based on their similarities in
  - gender,
  - ethnicity,
  - research fields,
  - affiliation,
  - whether authors have advisor-advisee relationships,
  - whether authors were coauthors in the past, and
  - whether authors share common coauthors in the past.

[21] Marcel Fafchamps, Marco J. Van der Leij, and Sanjeev Goyal. "Matching and network effects". *Journal of the European Economic Association* 8.1 (2010), pp. 203–231. ISSN: 1542-4766.

TABLE: Estimation results for the 2010-2012 sample with assortative matching.

		Homogeneous Spillovers	Heterogeneous Spillovers	Discounting # Coauthors
Output				
Spillover	(λ)	0.0604***	0.1103***	0.2501***
		(0.0207)	(0.0336)	(0.0654)
Cost	(ϕ)	0.3747***	0.3724***	0.3112***
		(0.0516)	(0.0256)	(0.0248)
Constant	(β_0)	-2.4953***	-2.2893***	-2.8615***
		(0.1710)	(0.1292)	(0.1911)
Log life-time citations	(β_1)	0.4836***	0.4566***	0.5543***
		(0.0228)	(0.0164)	(0.0270)
Decades after graduation	(β_2)	-0.3946***	-0.4019***	-0.5823***
		(0.0341)	(0.0243)	(0.0413)
Female	(β_3)	-0.5525***	-0.5228***	-0.3302***
		(0.0641)	(0.0525)	(0.0430)
NBER connection	(β_4)	0.5003***	0.5115***	0.4841***
		(0.0310)	(0.0318)	(0.0324)
Ivy League connection	(β_5)	0.3072***	0.2749***	0.2044***
		(0.0319)	(0.0284)	(0.0354)
Editor	(β_6)	0.2243***	0.1357***	0.1045***
		(0.0658)	(0.0443)	(0.0716)
Author effect	(ζ)	2.7577***	2.4732***	2.8271***
		(0.1290)	(0.0921)	(0.1172)
Project effect	(η)	3.5613***	2.5666***	2.6703***
		(0.3804)	(0.3841)	(0.3960)
Project variance	(σ_v^2)	74.7478***	78.9311***	77.5822***
		(1.7926)	(1.9171)	(1.9087)
Sample size		3,620	3,620	3,620

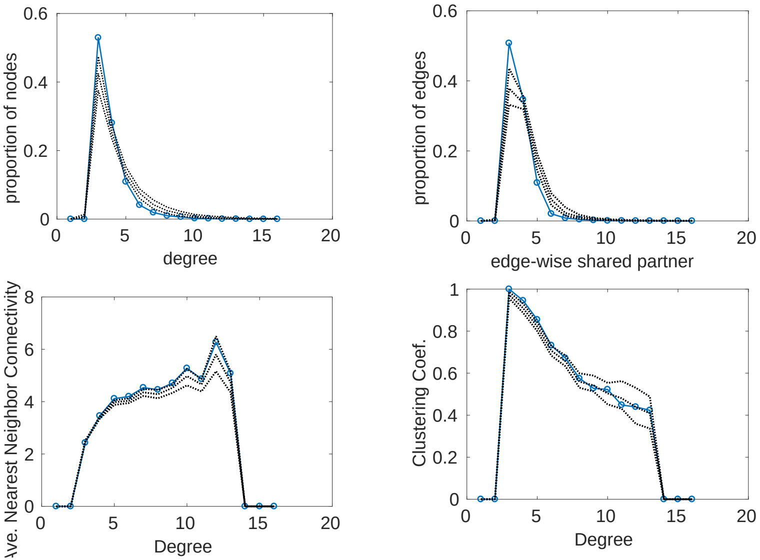
[a] Model (1): Assuming exogenous matching between authors and papers. Model (2): Assuming endogenous matching by Equation (13). The asterisks ***(**,*) indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.

TABLE: Estimation results for the 2010-2012 sample with assortative matching.

		Homogeneous Spillovers	Heterogeneous Spillovers	Discounting # Coauthors
Matching				
Constant	(γ_0)	-15.7636***	-15.0411***	-14.6548***
		(0.2146)	(0.2079)	(0.2015)
Same NEP	(γ_{11})	6.3917***	6.0590***	5.9113***
		(0.1245)	(0.1187)	(0.1199)
Ethnicity	(γ_{12})	5.2987***	5.0124***	4.7908***
		(0.1037)	(0.0997)	(0.0989)
Affiliation	(γ_{13})	5.3597***	5.1550***	4.9439***
		(0.2944)	(0.2845)	(0.2682)
Gender	(γ_{14})	2.5743***	2.3825***	2.3419***
		(0.1352)	(0.1490)	(0.1251)
Advisor-advisee	(γ_{15})	5.9521***	5.3956***	5.2407***
		(0.2112)	(0.2065)	(0.1917)
Past coauthors	(γ_{16})	6.1422***	5.9501***	5.8369***
		(0.1557)	(0.1230)	(0.1384)
Share common co-authors	(γ_{17})	10.1358***	9.5724***	9.2923***
		(0.1487)	(0.1282)	(0.1506)
Author effect	(γ_2)	4.4730***	4.0568***	4.3669***
		(0.1138)	(0.1104)	(0.1130)
Project effect	(γ_3)	-4.4184***	-3.7723***	-3.7262***
		(0.1129)	(0.0999)	(0.1187)
Sample size		3,620	3,620	3,620

[a] Model (1): Assuming exogenous matching between authors and papers. Model (2): Assuming endogenous matching by Equation (13). The asterisks ***(**,*) indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.

Goodness-of-Fit Statistics



Rankings for Individuals

Table: Ranking of the top-twenty five researchers from the 2010-2012 sample.

Rank	Name	Proj.	Citat.	RePEc Rank[a]	Output Loss[b]	Close.[c]	Betw.[c]	NEP Cites[d]	NEP IPR[e]	Organization
1	Van Reenen, John	20	6273	87	-1.91%	4.17	52.57	94.82	21.3983	London School of Economics
2	Alesina, Alberto	12	13625	39	-1.78%	4.17	29.47	94.86	18.7128	Harvard University
3	Ottaviano, Gianmarco	17	4302	220	-1.72%	4.14	39.49	91.69	14.4578	London School of Economics
4	Saez, Emmanuel	14	3930	314	-1.69%	4.61	4.64	91.70	11.1100	University of California-Berkeley
5	Reinhart, Carmen	12	18358	20	-1.60%	4.67	5.64	93.75	9.07441	Harvard University
6	Angrist, Joshua	6	8230	53	-1.60%	4.55	9.55	95.87	10.8803	Massachusetts Institute of Technology
7	List, John	27	7741	27	-1.59%	4.12	112.67	95.85	8.72953	University of Chicago
8	Nunn, Nathan	12	1495	656	-1.55%	4.88	0.54	90.52	11.9138	Harvard University
9	Bergemann, Dirk	32	1018	951	-1.54%	5.12	2.36	72.28	6.88862	Yale University
10	Pischke, Jorn-Steffen	9	2968	459	-1.54%	4.69	3.12	95.67	13.7850	London School of Economics
11	Rogoff, Kenneth	8	21001	8	-1.52%	4.42	10.04	94.81	12.9806	Harvard University
12	Melitz, Marc	9	6763	145	-1.47%	4.76	1.69	92.73	6.85287	Harvard University
13	Galor, Oded	12	7663	84	-1.46%	4.86	4.02	91.75	11.2015	Brown University
14	Wacziarg, Romain	8	2660	658	-1.44%	4.75	1.89	92.60	10.8025	University of California-Los Angeles
15	Bloom, Nicholas	12	4202	188	-1.36%	4.45	8.55	94.81	15.2667	Stanford University
16	Morris, Stephen	25	3414	284	-1.35%	4.47	11.07	87.55	6.56941	Princeton University
17	Wolfers, Justin	15	2786	607	-1.34%	4.71	3.64	93.69	18.9070	University of Michigan
18	Frankel, Jeffrey	41	10765	44	-1.34%	4.41	15.21	93.71	13.0174	Harvard University
19	Rasul, Imran	11	1447	906	-1.33%	4.57	5.60	85.55	17.5409	University College London
20	Borjas, George	8	6467	114	-1.32%	4.66	6.70	92.65	8.07620	Harvard University
21	Eichenbaum, Martin	6	10252	68	-1.29%	4.87	1.61	92.65	8.75977	Northwestern University
22	Black, Sandra	5	2813	563	-1.29%	4.77	2.45	93.68	9.05840	University of Texas-Austin
23	Lochner, Lance	12	2085	900	-1.27%	4.88	2.66	86.51	9.36363	University of Western Ontario
24	Basu, Susanto	3	2488	649	-1.22%	4.66	2.91	89.55	11.1390	Boston College
25	Demirguc-Kunt, Asli	13	9675	98	-1.18%	4.49	8.10	94.73	15.4880	World Bank Group

[a] The RePEc ranking is based on an aggregate of rankings by different criteria (cf. Zimmermann, 2013).
[b] The output loss for researcher i is computed as $\sum_{s=1}^{p} Y_s(G) - \sum_{s=1}^{p} Y_s(G^{-i})$ with the parameter estimates from Table 3.
[c] Betweenness centrality measures the fraction of all shortest paths in the network that contain a given node.
[d] NEP cites measures the breadth of citations across NEP fields.
[e] To gauge the degree of specialization of an author, we compute the inverse participation ratio (IPR) of the NEP fields in which the papers of an author are announced.

Rankings for Departments

Table: Ranking of the top-ten departments from the 2010-2012 sample.

Rank	Organization	Size	RePEc Rank[a]	Output Loss[b]
1	Department of Economics, Harvard University	23	1	-7.46%
2	Kennedy School of Government, Harvard University	14	16	-4.72%
3	Department of Economics, Princeton University	12	8	-4.28%
4	Economics Department, Massachusetts Institute of Technology	12	5	-3.29%
5	Centre for Economic Performance, London School of Economics	8	71	-3.19%
6	Economics Department, University of Michigan	16	31	-3.17%
7	Booth School of Business, University of Chicago	13	6	-2.79%
8	Department of Economics, University of California-Berkeley	10	10	-2.78%
9	Department of Economics, University of Pennsylvania	11	36	-2.76%
10	Economics Department, Yale University	10	19	-2.70%

[a] The RePEc ranking is based on an aggregate of rankings by different criteria (cf. Zimmermann, 2013).

[b] The output loss for department $\mathcal{D}$ is computed as $\sum_{s=1}^{p} Y_s(G) - \sum_{s=1}^{p} Y_s(G \backslash \mathcal{D})$ with the parameter estimates from Table 3. See also Equation (7).

Research Funding

- We compare our optimal research funding scheme r^* of Equation (11) using the parameter estimates with funding programs being implemented in the real world.[22]
- We use data on the funding amount, the receiving economics department and the principal investigators from the Economics Program of the *National Science Foundation (NSF)* in the U.S.[23]
- The National Bureau of Economic Research (NBER) received the largest amount of funds totalling to 95,058,724 U.S. dollars, followed by the University of Michigan with a total of 57,749,679 U.S. dollars.[24]

[22] Paula E. Stephan. *How economics shapes science*. Harvard University Press, 2012; Gianni De Frajay. "Optimal Public Funding for Research: A Theoretical Analysis". *RAND Journal of Economics* 47.3 (2016), pp. 498–528.

[23] See `https://www.nsf.gov/awardsearch/`.

[24] L. Drutman. "How the NSF allocates billions of federal dollars to top universities". *Sunlight foundation blog* (2012).

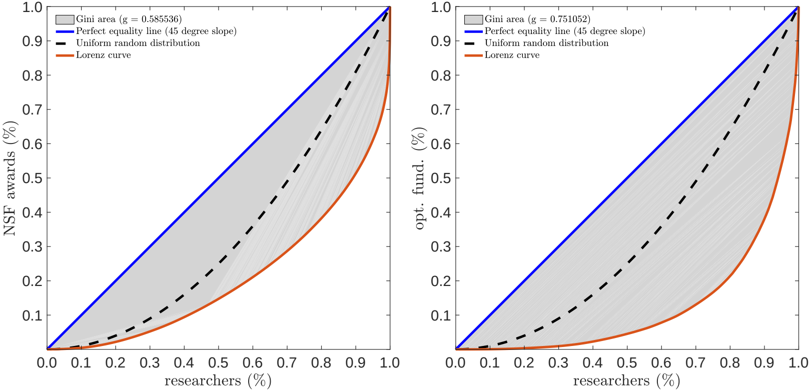


FIGURE: Lorenz curves for the total NSF awards (left panel) and the optimal network-based funding across authors (right panel). The concentration of funds towards the most productive researchers is even higher than for the NSF awards, with a Gini coefficient of $g = 0.59$ for the NSF awards and a coefficient of $g = 0.75$ for our network-based optimal funding policy.

Research Funding for Individuals

Table: Ranking of the optimal research funding for the top-twenty five researchers for the 2010-2012 sample.[a]

Name	Proj.	Deg.	Citat.	RePEc Rank[b]	Closen.[c]	Between.[c]	NEP Cites[d]	NEP IPR[e]	Organization	NSF [%]	Fund. [%][f]	Rank
Greg Kaplan	21	5	325	3261	5.0279	1.0236	11.0300	9.8561	Princeton University	0.0875	3.1251	1
Dirk Bergemann	36	5	1018	951	5.1203	2.3644	65.1200	6.8886	Yale University	0.0906	3.0984	2
Nicholas Bloom	13	4	4202	188	4.4500	8.5500	39.1200	15.2668	Stanford University	0.2982	2.7051	3
Olivier Coibion	11	3	765	1699	5.1402	0.7708	76.3600	5.1180	University of Texas-Austin	0.1017	2.5688	4
Fabrizio Perri	11	7	1909	738	4.9014	2.3939	65.1600	7.5378	Fed. Minneapolis	0.0414	2.4166	5
Stephen Morris	31	4	3414	284	4.4700	11.0700	42.1100	6.5694	Princeton University	0.2152	2.4116	6
Emmanuel Saez	14	7	3930	314	4.6100	4.6400	68.2100	11.1100	University of California-Berkeley	0.2786	2.3734	7
John List	29	12	7741	27	4.1200	112.6700	83.3500	8.7295	University of Chicago	0.0133	2.3509	8
Oded Galor	17	3	7663	84	4.8640	4.0200	37.0900	11.2016	Brown University	0.0822	2.3017	9
Sergio Rebelo	9	4	8043	127	4.9348	1.7689	13.0300	9.1398	Centre for Economic Policy Research	0.0890	2.2698	10
Craig Burnside	10	3	2700	578	5.1033	0.7259	2.0000	9.5135	Duke University	0.0426	2.2153	11
Yuriy Gorodnichenko	8	3	1940	495	4.5500	14.6800	71.2600	14.4722	University of California-Berkeley	0.0839	2.1810	12
Martin Eichenbaum	7	4	10252	68	4.8668	1.6054	89.6000	8.7598	Northwestern University	0.0500	1.9320	13
Vincenzo Quadrini	8	5	1460	1359	5.0292	0.5879	38.1100	9.5144	University of Southern California	0.1836	1.8019	14
Javier Bianchi	9	3	325	3654	5.4083	0.3712	72.2600	6.7236	Fed. Minneapolis	0.0418	1.7946	15
Joshua Angrist	6	3	8230	53	4.5500	9.5500	92.6100	10.8804	MIT	0.2597	1.7665	16
Andrei Levchenko	12	6	1081	1120	4.8565	2.4816	95.8100	8.4074	University of Michigan	0.0531	1.7374	17
Sandra Black	5	2	2813	563	4.7700	2.4471	31.0600	9.0584	University of Texas-Austin	0.0588	1.5930	18
Mark Huggett	8	4	1146	1245	5.5324	0.8559	30.0200	7.0588	Georgetown University	0.0128	1.5694	19
John Campbell	5	4	14782	11	4.5900	8.5600	27.0500	11.7769	Harvard University	0.0532	1.5349	20
Chad Syverson	6	4	1656	574	4.7300	4.3500	33.0700	13.8710	University of Chicago	0.0998	1.4443	21
Parag Pathak	3	4	1271	1130	4.8221	2.4116	36.1000	6.7827	NBER	0.2258	1.3842	22
Mikhail Golosov	12	9	1077	1025	4.7893	2.3778	3.0000	12.1462	Princeton University	0.0798	1.3701	23
Xavier Gabaix	8	2	3566	185	4.7818	2.5671	72.2500	19.1291	Harvard University	0.1378	1.3505	24
Aleh Tsyvinski	10	7	809	1388	4.8759	1.3933	78.3600	16.1578	Yale University	0.1550	1.2963	25

[a] We only consider the 236 researchers that are listed as principal investigators in the Economics Program of the National Science Foundation (NSF) in the U.S. from 1976 to 2016 and that can be identified in the RePEc database.
[b] The RePEc ranking is based on an aggregate of rankings by different criteria (cf. Zimmermann, 2013).
[c] See also Footnote c in Table 6.
[d] NEP cites measures the breadth of citations across NEP fields. See also Footnote d in Table 6.
[e] The inverse participation ratio (IPR) of the NEP fields meaures the degree of specialization of an author. See also Footnote e in Table 6.
[f] The total cost of funds, $\sum_{s=1}^{p} \delta_{is} r^* Y_s(\mathcal{G}, \mathbf{e}(r^*))$, of researcher i with the optimal research funding scheme r^* of Equation (11).

Research Funding for Departments

Table: Ranking of optimal research funding for the top-ten departments for the 2010-2012 sample.[a]

Institution	Size	NSF [%]	Funding [%][b]	Rank
Yale University	22	2.8771	8.3996	1
Princeton University	14	2.8250	8.1934	2
Harvard University	46	3.0338	6.7453	3
University of California-Berkeley	24	2.1543	5.9105	4
Federal Reserve Bank of Minneapolis	7	0.2578	5.7235	5
University of Chicago	30	2.6975	4.8246	6
Massachusetts Institute of Technology	18	1.7755	4.6533	7
University of Texas-Austin	12	0.3493	4.5853	8
Stanford University	21	4.0589	4.1962	9
University of Pennsylvania	22	3.0273	4.1644	10

[a] We only consider the 236 researchers that are listed as principal investigators in the Economics Program of the National Science Foundation (NSF) in the U.S. from 1976 to 2016 and that can be identified in the RePEc database.
[b] The total cost of funds, $\sum_{i \in \mathcal{D}} \sum_{s=1}^{p} \delta_{is} r^* Y_s(\mathcal{G}, \mathbf{e}(r^*))$, for each department $\mathcal{D}$ and researchers $i \in \mathcal{D}$ with the optimal research funding scheme r^* of Equation (11).

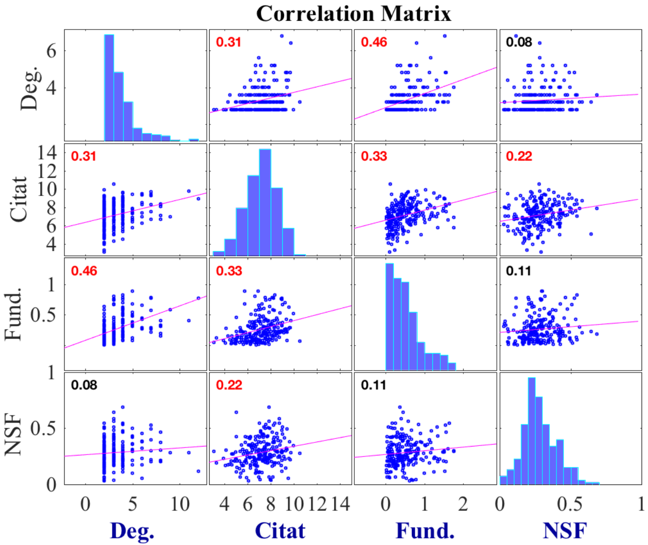


FIGURE: Pair correlation plot of the authors' degrees, citations, total NSF awards and the optimal funding policy. The Spearman correlation coefficients are shown for each scatter plot.

Conclusion

- We have analyzed the equilibrium efforts of authors involved in multiple, possibly overlapping projects.
- We bring our model to the data by analyzing the network of scientific coauthorships between economists registered in the RePEc author service.
- We rank the authors and their departments according to their contribution to aggregate research output, and thus provide the first ranking measure that is based on microeconomic foundations.
- Moreover, we analyze various funding instruments for individual researchers as well as their departments.
- We show that, because current research funding schemes do not take into account the availability of coauthorship network data, they are ill-designed to take advantage of the spillover effects generated in scientific knowledge production networks.

Future Work / Extensions

- We can allow the returns of an author from participating in a project to be split equally among the participants.[25]
- Instead of a convex cost, we can introduce a time constraint.[26]
- In work in progress we are extending our analysis to
  - the Framework Programs of the E.U. and
  - the research funding program of the Swiss National Science Foundation.

[25] Eugene Kandel and Edward P Lazear. "Peer pressure and partnerships". *Journal of political Economy* (1992), pp. 801–817; Matthew O. Jackson and Asher Wolinsky. "A Strategic Model of Social and Economic Networks". *Journal of Economic Theory* 71.1 (1996), pp. 44–74.

[26] Leonie Baumann. "Time allocation in friendship networks". *Available at SSRN 2533533* (2014); Hannu Salonen. "Equilibria and centrality in link formation games". *International Journal of Game Theory* 45.4 (2016), pp. 1133–1151.

Appendix

Individual author characteristics

1. Number of lifetime citations to all their works in their RePEc profile.
2. Number of times their works have been downloaded in the past 12 months from the RePEc services that report such statistics on LogEc (EconPapers, IDEAS, NEP, and Socionet).
3. Current RePEc ranking of the author. We use the aggregate ranking for the lifetime work.[27]
4. Current RePEc ranking for the main affiliation fo the author.
5. Year of the first publication recorded in the RePEc profile (article or paper).
6. Year of completion of terminal degree, as listed in the RePEc Genealogy.
7. Number of registered coauthors during career.
8. Dummy for editor of journal.
9. Dummy for NBER or CEPR affiliation.

[27] See `https://ideas.repec.org/top/top.person.all.html` for the top-ranked economists.

10. Dummy for terminal degree from an Ivy League institution.
11. Dummy for terminal degree obtained in the United States.
12. Dummy for main affiliation in the United States.
13. Gender as determined by a likelihood table using the first and possibly middle name. Uncertain matches were almost all resolved through internet search.
14. Ethnicity.
15. Closeness centrality measure.
16. Betweenness centrality measure.
17. Number of NEP fields in which author's work has been cited, to measure breadth of citations.
18. Inverse participation ratio for NEP fields of publications.
19. Fields of work, as determined by the NEP fields for which their working papers were selected for email dissemination.

Potential author pair characteristics

1. Co-authorship previous to the period under consideration.
2. Student-advisor relationship, as recorded in the RePEc Genealogy.
3. Joint alma mater of terminal decree as recorded in the RePEc Genealogy.
4. Joint affiliation, taken from the affiliations authors recorded in the RePEc Author Service. As authors may have multiple affiliations, we use two versions: one with only the main affiliation matching for the author-pair, the other where any of the affiliation matches.
5. Joint ethnicity.
6. Joint country of main affiliation.
7. Joint field of work. There are two ways we determine this, both based on the NEP fileds in which the authors published. For the first, we only consider the fields in which each author has written at least four papers or, for authors with less than 10 years of experience, a quarter of all papers announced through NEP. A match is called if at least one field coincides in the author pair. For the second, we consider for each author the proportion of papers in each fields, and then compute a score by multiplying the vectors of the authors across all fields.

Paper characteristics

1. Number of citations for all versions of the paper.
2. Same, but weighted simple impact factors, as listed on IDEAS.
3. Same, but weighted recursive impact factors, as listed on IDEAS.
4. Same, but weighted discounted impact factors, as listed on IDEAS.
5. Same, but weighted recursive discounted impact factors, as listed on IDEAS.
6. Same, but weighted simple discounted impact factors, as listed on IDEAS.
7. If published, the journal's simple impact factor, as listed on IDEAS.
8. If published, the journal's recursive impact factor, as listed on IDEAS.
9. If published, the journal's H-index, as listed on IDEAS.
10. The number of downloads in the last 12 months, as provided by LogEc.
11. The number of authors.

12. The average number of works across authors of this paper.
13. Same, weighted by simple impact factors.
14. Same, weighted by recursive impact factors.
15. The average number of citations across authors of this paper.
16. Same, weighted by simple impact factors.
17. Same, weighted by recursive impact factors.
18. The average number of citations across authors of this paper, each citation also divided by the number of authors of the cited paper.
19. Same, weighted by simple impact factors.
20. Same, weighted by recursive impact factors.
21. Number of references in the paper that have been matched with other items in RePEc.
22. Same, weighted by simple impact factors.
23. Same, weighted by recursive impact factors.
24. Year of publication in a journal.
25. Dummy if at least one author is editor.
26. Dummy if authors have main affiliations in different countries.

Exogenous vs. Endogenous Networks

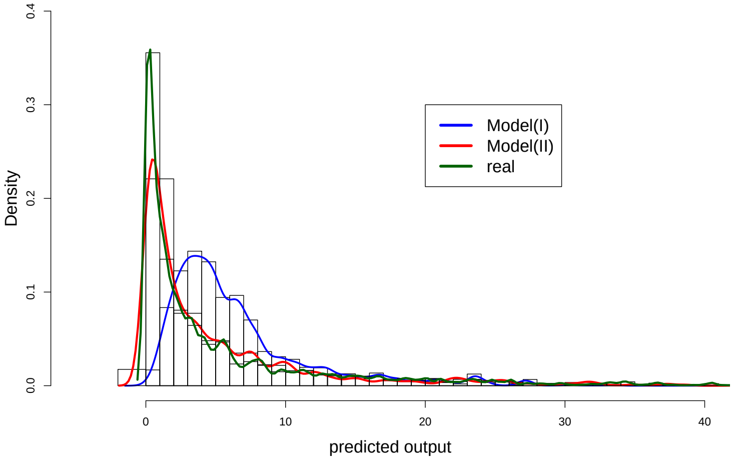


Figure: Distributions of paper qualities with an exogenous coauthor network (Model I) and endogenous author-project matching (Model II).