Superstar Economists: Coauthorship networks and research output

Chih-Sheng Hsieh, Michael D. König, Xiaodong Liu, Christian Zimmermann

Abstract
We study the impact of research collaborations in coauthorship networks on total research output. Through the links in the collaboration network researchers create spillovers not only to their direct coauthors but also to researchers indirectly linked to them. We characterize the interior equilibrium when agents spend effort in multiple, possibly overlapping projects, and there are interaction effects in the cost of effort. 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, and compare them to the economics funding program by the National Science Foundation. Our results indicate that, because current 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.

Key words: coauthor networks, scientific collaboration, spillovers, key player, research funding, economics of science

JEL: C72, D85, D43, L14, Z13

1. Introduction

We build a micro-founded model for the output produced in scientific co-authorship networks that incorporates and generalizes previous ones in the literature [cf. e.g. Ballester et al., 2006; Cabrales et al., 2010; Jackson and Wolinsky, 1996]. However, 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 to this model then allows us to study the impact of individual researchers on total research output (ranking), and the optimal design of research funding programs.

---

*We would like to thank Francis Bloch, Pietro Biroli, Ralph Ossa, Hannes Schwandt and Fabrizio Zilibotti for the helpful comments. Moreover, we would like to thank Adrian Etter and Marc Biedermann for the excellent research assistance. Michael D. König acknowledges financial support from Swiss National Science Foundation through research grant PZ00P1_154957 / 1. The views expressed are those of the individual authors and do not necessarily reflect official positions of the Federal Reserve Bank of St. Louis, the Federal Reserve System, or the Board of Governors.

Email addresses: csheih@cuhk.edu.hk (Chih-Sheng Hsieh), michael.koenig@econ.uzh.ch (Michael D. König), xiaodong.liu@colorado.edu (Xiaodong Liu), zimmermann@stlouisfed.org (Christian Zimmermann)
We further provide a novel estimation framework in which agents can participate in many potentially overlapping projects, and this 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 [cf. e.g. Chandrasekhar and Jackson, 2012], 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.¹

We then develop a novel ranking measure for economists and their departments which is derived from economic micro-foundations that takes explicitly spillovers between collaborating economists into account. Our ranking quantifies the endogenous decline in research output due to the removal of an economist from the coauthorship network [cf. Ballester et al., 2006; König et al., 2014], and allows us to determine “key players” [cf. Zenou, 2015], or “superstar” economists [cf. Azoulay et al., 2010; Waldinger, 2010, 2012].² Taking into account the endogenous effort made by the authors, and the spillovers generated across them in the coauthorship network, we find that the highest ranked authors are not necessarily the ones with the largest number of citations, or coincide with other author ranking measures used in the literature. However, this discrepancy is not surprising, as traditional rankings are typically not derived from microeconomic foundations, and typically do not take into account the spillover effects generated in scientific knowledge production networks.

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 [see also Stephan, 1996, 2012]. We analyze various funding instruments (targeted versus non-discriminatory) for individual researchers, and study how the funds to different researchers impact aggregate scientific output [cf. König et al., 2014]. We then aggregate researchers by their research institutions and departments, and compute the optimal funding for these institutions [cf. Aghion et al., 2010]. A comparison of our optimal funding policy with the research funding of the economics program of the National Science Foundation (NSF) indicates that there are significant differences, both at the individual and the departmental levels. In particular, we find that our optimal funding policy is significantly positively correlated with the degree (number of coauthors) and the number of lifetime citations of an author. In contrast, the NSF awards are not correlated with the degree and positively but not significantly correlated with the optimal funding policy. This highlights the importance of the coauthorship network in determining the

¹When two authors claim the same paper in the RePEc digital library, they are coauthors, and the relationship of coauthorship creates an undirected network between them. 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. See http://repec.org/ for a general description of the RePEc database.

²Note that the effect of hiring superstar scientists on the profitability of firms has been studied in Hess and Rothaermel [2011]; Lacetera et al. [2004]; Rothaermel and Hess [2007]. In particular, Rothaermel and Hess [2007] define star scientists as researchers who had both published and been cited at a rate of three standard deviations above the mean. In contrast, our measure of star scientists takes into account the spillover effects of one scientist on others in a collaboration network.
optimal funding policy, while it does not seem to have an effect on the research funding program by the NSF.

There exists a growing literature, both empirical and theoretical, on the formation and consequences of coauthorship networks. On the empirical side, the structural features of scientific collaboration networks have been analyzed in Goyal et al. [2006], Newman [2001a, 2004, 2001b,c,d] and König [2011]. Fafchamps et al. [2010] study predictors for the establishment of scientific collaborations, and Ductor [2014]; Ductor et al. [2014] study how these collaborations affect research output of individual authors. At an aggregate level, Bosquet and Combes [2013] estimate the effect of department size on its research output. Different to these works, we take a structural approach by introducing a production function for the output produced in a collaborative research project (paper), and endogenously determine the unobserved research effort [cf Ballester et al., 2006; Cabrales et al., 2010; Calvó-Armengol et al., 2009]. Moreover, we develop a micro-founded ranking measure of authors and their departments [cf. Azoulay et al., 2010; Liu et al., 2011; Waldinger, 2010, 2012],[3] and investigate optimal research funding policies [cf. De Frajaj, 2016; König et al., 2014; Stephan, 2012].

Our paper is further related to the recent theoretical contributions by Baumann [2014] and Salonen [2016], where agents choose time to invest into bilateral relationships. Our model extends the set-ups considered in these papers to allow for investments into multiple projects involving more than two agents. Moreover, in a related paper Bimpikis et al. [2014] analyze firms competing in quantities à la Cournot across different markets with a similar linear-quadratic payoff specification, and allow firms to choose endogenously the quantities sold to each market. Different to these authors, the efforts invested by the agents in different projects in our model are strategic complements, and not substitutes as in their papers.

Our paper is organized as follows. The scientific production function is introduced in Section 2, while agents’ payoffs are discussed in Section 3. The policy relevance of our model is illustrated in Section 4, where in Section 4.1 we investigate the impact of the removal of an author from the network, while in Section 4.2 we analyze optimal research funding schemes that take into account the spillovers generated across collaborating authors in the network. The empirical implications of the model are discussed in Section 5. The data used for this study is described in Section 5.1, and our econometric methodology is explained in Section 5.2. The matching process of authors and projects is introduced in Section 5.3, a Bayesian estimation method is discussed in Section 5.4 and estimation results are given in Section 5.5. The empirical key player analysis (both at the author and the department level) is then provided in Section 6. Section 7 provides the optimal research funding policy and compares it with the economics funding program by the National Science Foundation (NSF). Finally, Section 8 concludes. Appendix A provides additional examples that illustrate the Nash equilibrium effort levels. Alternative payoff specifications are discussed in Appendices B and C, while the proofs are relegated to

---

3There is also a large literature on how to rank authors/departments according to their productivity measured by citations. See for example Perry and Reny [2016], Palacios-Huerta and Volij [2004], Zimmermann [2013] and Labrano et al. [2009].
Appendix D. More detailed information about the data can be found in Appendix E and some relevant technical material can be found in Appendix F.

2. Production Function

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

\[
Y_s(G, e_s) = \sum_{i \in N_s} \alpha_i e_{is} + \frac{\beta}{2} \sum_{i \in N_s} \sum_{j \in N_s \setminus \{i\}} e_{is} e_{js} = \sum_{i \in N_s} e_{is} \left( \alpha_i + \frac{\beta}{2} \sum_{j \in N_s \setminus \{i\}} e_{js} \right),
\]

where \( e_{is} \geq 0 \) is the research effort of agent \( i \) in project \( s \), \( N_s \subseteq N \) is the set of agents participating in project \( s \), \( e_s \) is the stacked vector of authors’ effort levels participating in project \( s \), \( \alpha_i \geq 0 \) is the ability/skill of researcher \( i \) and \( \beta \) is a spillover parameter from complementarities between the research efforts of coauthors. If efforts are measured in logs then \( Y_s \) corresponds to a translog production function [cf. Christensen et al., 1973, 1975]. The translog production function can be viewed as an exact production function, a second order Taylor approximation to a more general production function or a second order approximation to a CES production function, and has been used for example to analyze production in teams [cf. Adams, 2006].

3. Payoffs

We assume that the payoff of author \( i \) is given by \( \pi_i(G, e) = \sum_{s=1}^{p} Y_s(G, e) \delta_{is} - c_i \), where \( \delta_{is} \in \{0, 1\} \) indicates whether author \( i \) is participating in project \( s \) and \( c_i \geq 0 \) is the cost of \( i \). In the following we consider a cost \( c_i \) given by the quadratic form

\[
c_i(G, e) = \frac{1}{2} \sum_{s,s'=1}^{p} \phi_{s,s'} e_{is} e_{is'} \delta_{is} \delta_{is'} = \frac{1}{2} \tilde{e}_i ^\top \phi \tilde{e}_i,
\]

where \( \phi_{s,s'} = \phi_{s',s} \), \( \tilde{e}_i = (\tilde{e}_{i1}, \ldots, \tilde{e}_{ip}) ^\top \) and \( \tilde{e}_{is} = e_{is} \delta_{is} \). This cost is convex if and only if the \( p \times p \) matrix \( \phi \) with elements \( \phi_{s,s'} \) is positive definite. The case of a quadratic cost includes the case of a convex total cost as a special case when \( \phi_{s,s'} = \gamma \), and the case of a convex separable cost discussed in Appendix C when \( \phi_{s,s'} = \gamma \delta_{s,s'} \). The introduction of a quadratic cost with substitutes or complements, depending on the sign of the parameters \( \phi_{s,s'} \), is similar to the model analyzed in Cohen-Cole et al. [2012]. A theoretical model with a similar specification but allowing for only two activities is studied in Belhaj and Deroïan [2014], and an empirical analysis is provided in Liu [2014]. Further, a convex separable cost can be similarly found in the model studied in Adams [2006].

---

4A related specification, however, without allowing agents to spend effort across different projects, can be found in Ballester et al. [2006] and Cabrales et al. [2010].

5Table 7 gives an overview of possible extensions and alternative specifications.
The payoff of agent $i$ is then given by

$$\pi_i(G, e) = \sum_{s=1}^{p} Y_s(G, e_s) \eta_{is} - \frac{1}{2} \sum_{s,s'=1}^{p} \phi_{s,s'} e_{is} e_{is'} \delta_{is} \delta_{is'}$$

$$= \sum_{s=1}^{p} \left( \sum_{j \in N_s} e_{js} \left( \alpha_j + \frac{\beta}{2} \sum_{k \in N_i \setminus \{j\}} e_{ks} \right) \right) \eta_{is} - \frac{1}{2} \sum_{s,s'=1}^{p} \phi_{s,s'} e_{is} e_{is'} \delta_{is} \delta_{is'}, \tag{3}$$

where $n_s = |N_s|$ is the number of agents participating in project $s$, and we have that $n_s = \sum_{i=1}^{n} \eta_{is}$ with $\eta_{is} \in \{0, 1\}$ indicating whether $i$ is participating in project $s$.

The following proposition provides a complete equilibrium characterization of the authors’ effort levels across multiple projects.

**Proposition 1.** Let the payoff function for each agent $i = 1, \ldots, n$ be given by Equation (3) and assume that

$$\phi_{ss'} = \begin{cases} \gamma, & \text{if } s' = s, \\ \rho, & \text{otherwise}. \end{cases}$$

Denote by

$$\varphi_{is} = \frac{\rho \beta \eta_{is}}{(\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1))},$$

$$\mu_s(\alpha) = \sum_{i=1}^{n} \frac{\rho \alpha_i d_i \eta_{is}}{(\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1))} - \frac{1}{\beta(n_s - 1) + \rho - \gamma} \sum_{i=1}^{n} \alpha_i \eta_{is},$$

$$\omega_{ss'} = \sum_{i=1}^{n} \varphi_{is} \eta_{is'}.$$

Further, let $\Omega = (\omega_{ss'})_{1 \leq s, s' \leq p}$, assume that the matrix $I_p - \Omega$ is invertible, and define by $\epsilon = (I_p - \Omega)^{-1} \mu(\alpha)$. Then, for $\beta$ small enough, the unique interior Nash equilibrium effort levels are given by

$$e_{is} = \frac{1}{\beta + \gamma - \rho} \left[ \beta \epsilon_s + \alpha_i - \frac{\rho}{\beta + \gamma + \rho(d_i - 1)} \left( \sum_{s'=1}^{p} \eta_{is'} \epsilon_{s'} + \alpha_i d_i \right) \right], \tag{4}$$

if $\eta_{is} = 1$ for each agent $i = 1, \ldots, n$ and each project $s = 1, \ldots, p$. Further, the total effort spent in project $s$ is given by $\epsilon_s$.

---

See the sufficient conditions derived in Appendix F to guarantee existence and uniqueness.
Inserting Equation (4) into the production function from Equation (1) gives

\[
Y_s(G) = \sum_{i \in \mathcal{N}_s} \alpha_i e_{is} + \beta \sum_{i \in \mathcal{N}_s} \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{is} e_{js} = \sum_{i \in \mathcal{N}_s} e_{is} \left( \alpha_i + \beta \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js} \right)
\]

\[
\times \left\{ \alpha_i + \beta \sum_{j \in \mathcal{N}_s \setminus \{i\}} \frac{\delta_{js}}{\beta + \gamma - \rho} \left[ \beta e_{is} + \alpha_j - \frac{\rho}{\beta + \gamma + \rho(d_j-1)} \left( \sum_{s'=1}^{p} \delta_{js's} e_{s'} + \alpha_{j} d_j \right) \right] \right\}.
\]

(5)

An alternative compact form of the equilibrium effort levels is derived in the following proposition.

**Proposition 2.** Let the payoff function for each agent \( i = 1, \ldots, n \) be given by Equation (3), assume that

\[
\phi_{ss'} = \begin{cases} 
\gamma, & \text{if } s' = s, \\
\rho, & \text{otherwise.}
\end{cases}
\]

and let \( \Gamma \) be the symmetric \((n \times p) \times (n \times p)\) matrix with elements

\[
\Gamma_{is,jk} = \begin{cases} 
\rho \delta_{is} \delta_{jk}, & \text{if } i = j, s \neq k, \\
-\beta \delta_{is} \delta_{jk}, & \text{if } i \neq j, s = k, \\
0, & \text{otherwise,}
\end{cases}
\]

(6)

with zero diagonal for \( i, j = 1, \ldots, n \), and \( s, k = 1, \ldots, p \). Further let \( \delta \) be an \((n \times p)\) stacked vector with elements \( \delta_{is} \), \( \alpha \) an \((n \times p)\) stacked vector with elements \( \alpha_{is} = \alpha_i \delta_{is} \) and \( e \) an \((n \times p)\) stacked vector of agent-project effort levels, \( e_{is} \). Then if the matrix \( \gamma \text{ diag}(\delta) + \Gamma \) is invertible,\(^6\) the equilibrium effort levels are given by

\[
e = (\gamma \text{ diag}(\delta) + \Gamma)^{-1} \alpha.
\]

Observe that the matrix \( \Gamma \) represents a weighted matrix of the line graph\(^7\) \( L(G) \) of the bipartite collaboration network \( G \), where each link between nodes sharing a project has weight \( -\beta \), and each link between nodes sharing an author has weight \( \rho \). An example can be found in Figure 1.\(^8\)

We will illustrate the equilibrium characterization of Propositions 1 and 2 in the following example (see also Figure 1). Further examples can be found in Appendix A.

---

\(^6\)Given a graph \( G \), its line graph \( L(G) \) is a graph such that each node of \( L(G) \) represents an edge of \( G \), and two nodes of \( L(G) \) are adjacent if and only if their corresponding edges share a common endpoint in \( G \) [cf. e.g. West, 2001].

\(^7\)This representation is similar to equilibrium characterization considered in Bimpikis et al. [2014]. Further examples can be found in Figures A.1 and A.3 in Appendix A.
Example 1. Consider a network with 2 projects and 3 agents, where in the first project agents 1 and 2 are collaborating and in the second project agents 1 and 3 are collaborating. An illustration can be found in Figure 1. The payoffs of the agents are given by

\[
\begin{align*}
\pi_1 &= e_{11} \left( \alpha_1 + \frac{\beta}{2} e_{21} \right) + e_{21} \left( \alpha_2 + \frac{\beta}{2} e_{11} \right) + e_{12} \left( \alpha_1 + \frac{\beta}{2} e_{32} \right) + e_{32} \left( \alpha_3 + \frac{\beta}{2} e_{12} \right) \\
&\quad - \frac{\gamma}{2} e_{11} - \frac{\gamma}{2} e_{12} - \rho e_{11} e_{12} \\
\pi_2 &= e_{11} \left( \alpha_1 + \frac{\beta}{2} e_{21} \right) + e_{21} \left( \alpha_2 + \frac{\beta}{2} e_{11} \right) - \frac{\gamma}{2} e_{21} \\
\pi_3 &= e_{32} \left( \alpha_3 + \frac{\beta}{2} e_{12} \right) + e_{32} \left( \alpha_3 + \frac{\beta}{2} e_{12} \right) - \frac{\gamma}{2} e_{32}.
\end{align*}
\]

The first order conditions are given by

\[
\begin{align*}
\frac{\partial \pi_1}{\partial e_{11}} &= \alpha_1 + e_{21} \beta - e_{11} \gamma - e_{12} \rho = 0 \\
\frac{\partial \pi_1}{\partial e_{12}} &= \alpha_1 + e_{32} \beta - e_{12} \gamma - e_{11} \rho = 0 \\
\frac{\partial \pi_2}{\partial e_{21}} &= \alpha_2 + e_{11} \beta - e_{21} \gamma = 0 \\
\frac{\partial \pi_3}{\partial e_{32}} &= \alpha_3 + e_{12} \beta - e_{32} \gamma = 0.
\end{align*}
\]
Solving this system of equations directly yields

\[
\begin{align*}
e_{11} &= \frac{(\beta - \gamma)(\beta + \gamma)(\alpha_2 \beta + \alpha_1 \gamma) + \gamma(\alpha_3 \beta + \alpha_1 \gamma)}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2}, \\
e_{12} &= \frac{(\beta - \gamma)(\beta + \gamma)(\alpha_3 \beta + \alpha_1 \gamma) + \gamma(\alpha_2 \beta + \alpha_1 \gamma)}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2}, \\
e_{21} &= \frac{(\beta - \gamma)(\beta + \gamma)(\alpha_1 \beta + \alpha_2 \gamma) + \beta(\alpha_3 \beta + \alpha_1 \gamma)}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2}, \\
e_{32} &= \frac{(\beta - \gamma)(\beta + \gamma)(\alpha_1 \beta + \alpha_3 \gamma) + \beta(\alpha_2 \beta + \alpha_1 \gamma)}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2}.
\end{align*}
\]

Next, we compute the above equilibrium effort levels using the equilibrium characterization in Equation (4). Note that \(d = (d_i)_{1 \leq i \leq 3} = (2, 1, 1)^T\), \(n = (n_s)_{1 \leq s \leq 2} = (2, 2)^T\),

\[
\delta = (\delta_{is})_{1 \leq i \leq 3, 1 \leq s \leq 2} = \begin{pmatrix} 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},
\]

and

\[
\varphi = (\varphi_{is})_{1 \leq i \leq 3, 1 \leq s \leq 2} = \begin{pmatrix} \beta_p & \beta_p \\ \beta_p & 0 \\ 0 & \beta_p \end{pmatrix}.
\]

Further, we have that \(\alpha = (\alpha_1, \alpha_2, \alpha_3)^T\) and

\[
\mu(\alpha) = \begin{pmatrix} \alpha_2 \left(-1 + \frac{\rho}{\beta - \gamma + \rho}\right) + \alpha_1 \left(-1 + \frac{2 \rho}{\beta + \gamma + \rho}\right) \\ \alpha_3 \left(-1 + \frac{\rho}{\beta - \gamma + \rho}\right) + \alpha_1 \left(-1 + \frac{2 \rho}{\beta + \gamma + \rho}\right) \\ \frac{1}{\beta - \gamma + \rho} \end{pmatrix}.
\]

Next, we have that

\[
\Omega = \begin{pmatrix} \frac{\beta_p(2(\beta + \gamma) + \rho)}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} & \frac{\beta_p}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} \\ \frac{\beta_p(\beta + \gamma + \rho)}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} & \frac{\beta_p}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} \\ \frac{\beta_p(2(\beta + \gamma) + \rho)}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} & \frac{\beta_p}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} \end{pmatrix},
\]

and hence

\[
\epsilon = (I_2 - \Omega)^{-1} \mu(\alpha) = \begin{pmatrix} \frac{(\alpha_1 + \alpha_2)(\beta - \gamma)(\beta + \gamma)^2 + (\beta + \gamma)(\alpha_3 \beta + \alpha_1 \gamma) + \alpha_2 \gamma \rho^2}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2} \\ \frac{(\alpha_1 + \alpha_3)(\beta - \gamma)(\beta + \gamma)^2 + (\beta + \gamma)(\alpha_3 \beta + \alpha_1 \gamma) + \alpha_3 \gamma \rho^2}{\gamma^2 \rho^2 - (\beta^2 - \gamma^2)^2} \\ 0 \end{pmatrix}.
\]

Inserting the above expressions into Equation (4) yields exactly the equilibrium effort levels of Equation (9).

We next compute the equilibrium following Proposition 2. The matrix \(\Gamma\) with elements \(\Gamma_{is,jk}\)
from Equation (6) can be written as follows

\[
\begin{pmatrix}
\gamma & 0 & 0 & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 & 0 & 0 \\
0 & 0 & \gamma & 0 & 0 & 0 \\
0 & 0 & 0 & \gamma & 0 & 0 \\
0 & 0 & 0 & 0 & \gamma \\
\end{pmatrix}

\begin{pmatrix}
0 & \rho & -\beta & 0 & 0 & 0 \\
\rho & 0 & 0 & 0 & 0 & -\beta \\
-\beta & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\beta & 0 & 0 & 0 & 0 \\
\end{pmatrix}

\begin{pmatrix}
\epsilon_{11} \\
\epsilon_{12} \\
\epsilon_{21} \\
\epsilon_{22} \\
\epsilon_{31} \\
\epsilon_{32} \\
\end{pmatrix}

= \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\end{pmatrix}
\]

The FOCs can be written as\(^9\)

\[
\begin{pmatrix}
\gamma_{e_{11}} + \rho e_{12} - \beta e_{21} \\
\gamma e_{12} + \rho e_{11} - \beta e_{32} \\
\gamma e_{21} - \beta e_{11} \\
0 \\
0 \\
\gamma e_{32} - \beta e_{12} \\
\end{pmatrix}

= \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\end{pmatrix}
\]

from which we find that

However, these are exactly the FOC from Equation (8).

An illustration of the equilibrium effort levels for agent 1 with \(\alpha_1 = 0.2, \alpha_2 = 0.1, \alpha_3 = 0.9,\) \(\gamma = 1\) and two different values of \(\rho = 0.05\) and \(\rho = 0.25\) is shown in Figure 2. We observe that with increasing values of \(\beta\) the effort spent by agent 1 on project 2 is increasing more than the effort spent on project 1. The reason is that with increasing \(\beta\) the complementarity effects between efforts of collaborating agents become stronger, and this effect is more pronounced for the collaboration of agent 1 with the more productive agent 3, than with the less productive agent 2. Moreover, when the cost parameter \(\rho\) is high enough, then agent 1 may even spend less effort in equilibrium in the project with agent 1 for higher values of \(\beta\) than for low values of \(\beta\), indicating congestion and substitution effects across projects.

4. Policy Implications

In the following we analyze the importance of authors and their departments (cf. Section 4.1), and we investigate how research funds should optimally be allocated to them (cf. Section 4.2).

4.1. Superstars, Key Players and Rankings

In this section we analyze the impact of the removal of individual authors from the coauthorship network on overall scientific output [cf. e.g. Waldinger, 2010, 2012]. The author whose removal

---

\(^9\)See also Equation (63) in the proof of Proposition 2 in Appendix D.
Figure 2: Equilibrium effort levels for agent 1 with $\alpha_1 = 0.2, \alpha_2 = 0.1, \alpha_3 = 0.9, \rho = 0.05$ (left panel), $\rho = 0.25$ (right panel) and $\gamma = 1$. The dashed lines indicate the effort level for $\beta = 0$.

would result in the greatest loss is termed the “key author” [Zenou, 2015] or “superstar” [Azoulay et al., 2010]. More formally, the key author is defined by

$$i^* \equiv \arg\max_{i \in \mathcal{N}} \left\{ \sum_{s=1}^{p} Y_s(G) - \sum_{s=1}^{p} Y_s(G^{-i}) \right\}. \quad (10)$$

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

$$\mathcal{D}^* \equiv \arg\max_{\mathcal{D} \subset \mathcal{N}} \left\{ \sum_{s=1}^{p} Y_s(G) - \sum_{s=1}^{p} Y_s(G \setminus \mathcal{D}) \right\}. \quad (11)$$

### 4.2. Research Funding

In the following we consider a two-stage game: In the first stage, the planner announces the research funding scheme $z \in \mathbb{R}_+^n$ that the authors should receive, and in the second stage the authors choose their research efforts, given $z$. The optimal funding profile $z^*$ can then be found by backward induction.\textsuperscript{11} Aggregating the individual funds to the department level also allows us to determine the optimal research funding for departments. For a general discussion of funding of academic research see Stephan [1996, 2012].

We first solve the second stage of the game. We assume that agent $i \in \mathcal{N}$ receives research funding, $z_i \geq 0$, proportional to the output she generates. When the cost is convex in the total

\textsuperscript{10}Note that our model can also be used to measure the potential loss (gain) on research output of a department due to a faculty member leaving (joining) one department for another. This could guide the academic wage bargaining process when professors get an offer from a competing university.

\textsuperscript{11}A similar planner’s problem in the context of subsidies to R&D collaborating firms has been analyzed in König et al. [2014].
effort spent then we get

\[
\pi_i(G, \mathbf{e}, \mathbf{z}) = \sum_{s=1}^{p} Y_s(G, e_s) \delta_{is} - \frac{1}{2} \sum_{s,s'=1}^{p} \phi_{s,s'} e_{is} e_{is'} \delta_{is} \delta_{is'} + \sum_{s=1}^{p} Y_s(G, e_s) z_i \delta_{is} \quad (12)
\]

Assuming that

\[
\phi_{s,s'} = \begin{cases} 
\gamma, & \text{if } s' = s, \\
\rho, & \text{otherwise,}
\end{cases}
\]

and denoting by \( \tilde{z}_i \equiv 1 + z_i \geq 1 \), we can write the payoff as follows

\[
\pi_i(G, \mathbf{e}, \tilde{\mathbf{z}}) = \sum_{s=1}^{p} \tilde{z}_i \left( \sum_{j \in \mathcal{N}_s} e_{js} \left( \alpha_j + \frac{\beta}{2} \sum_{k \in \mathcal{N}_s \setminus \{j\}} e_{ks} \right) \right) \delta_{is} - \frac{1}{2} \gamma \sum_{s=1}^{p} e_{is}^2 \delta_{is} - \rho \sum_{s=1}^{p} \sum_{s' \neq s} e_{is} e_{is'} \delta_{is} \delta_{is'}. \quad (13)
\]

The FOC w.r.t. \( e_{is} \) are given by

\[
\frac{\partial \pi_i(G, \mathbf{e}, \tilde{\mathbf{z}})}{\partial e_{is}} = \tilde{z}_i \left( \alpha_i + \frac{\beta}{2} \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js} \right) \delta_{is} - \gamma e_{is} \delta_{is} = 0.
\]

Solving for \( e_{is} \) yields

\[
e_{is} = \frac{\alpha_i \tilde{z}_i}{\gamma} + \frac{\beta}{\gamma} \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js}. \quad (14)
\]

This equation can be solved numerically using a fixed point algorithm [Nocedal and Wright, 2006]. Next, making the simplifying assumption that all researchers obtain the same funding per output produced, i.e. we set \( \tilde{z}_i = \tilde{z} \) for all \( i = 1, \ldots, n \), we then have that the payoff of agent \( i \) is given by

\[
\pi_i(G, \mathbf{e}, \tilde{\mathbf{z}}) = \sum_{s=1}^{p} \tilde{z}_i \left( \sum_{j \in \mathcal{N}_s} e_{js} \left( \tilde{\alpha}_j + \frac{\tilde{\beta}}{2} \sum_{k \in \mathcal{N}_s \setminus \{j\}} e_{ks} \right) \right) \delta_{is} - \frac{1}{2} \gamma \sum_{s=1}^{p} e_{is}^2 \delta_{is} - \rho \sum_{s=1}^{p} \sum_{s' \neq s} e_{is} e_{is'} \delta_{is} \delta_{is'}. \quad (15)
\]

If we denote by \( \tilde{\alpha}_i \equiv \frac{\alpha_i}{\tilde{z}} \) and \( \tilde{\beta} \equiv \frac{\beta}{\tilde{z}} \) then we get

\[
\pi_i(G, \mathbf{e}, \tilde{\mathbf{z}}) = \sum_{s=1}^{p} \left( \sum_{j \in \mathcal{N}_s} e_{js} \left( \tilde{\alpha}_j + \frac{\tilde{\beta}}{2} \sum_{k \in \mathcal{N}_s \setminus \{j\}} e_{ks} \right) \right) \delta_{is} - \frac{1}{2} \gamma \sum_{s=1}^{p} e_{is}^2 \delta_{is} - \rho \sum_{s=1}^{p} \sum_{s' \neq s} e_{is} e_{is'} \delta_{is} \delta_{is'}. \quad (16)
\]

This is the same functional form for the agent’s payoffs as we have encountered already in
Equation (3), and hence we obtain the equilibrium effort levels as in Proposition 1 given by

\[ e_{is} = \frac{1}{\beta + \gamma - \rho} \left[ \beta \epsilon_s + \hat{\alpha}_i - \frac{\rho}{\beta + \gamma + \rho (d_i - 1)} \left( \sum_{s' = 1}^{p} \delta_{is'} \epsilon_{s'} + \hat{\alpha}_i d_i \right) \right], \]

(17)

with \( \varphi_{is}, \mu_s(\alpha), \omega_{ss'}, \Omega \) and \( \epsilon \) as in Proposition 1.

In order to solve the first stage of the game, the planner has to solve the following problem

\[
\begin{align*}
    z^* &= \arg\max_{z \in \mathbb{R}_{\geq 0}} \sum_{s=1}^{p} Y_s(G, z) - \sum_{i=1}^{n} \sum_{s=1}^{p} \delta_{is} (z - 1) Y_s(G, z) - 1 = \arg\max_{z \in \mathbb{R}_{\geq 0}} \sum_{i=1}^{n} \sum_{s=1}^{p} \left( \frac{1}{n} - z \delta_{is} \right) Y_s(G, z - 1),
\end{align*}
\]

(18)

where \( Y_s(G, z) \) is the output of project \( s \) from Equation (1) with effort levels given by Equation (17). The general case of heterogeneous research funding, where \( z_i \neq z_j \), can be solved numerically with effort levels given by Equation (14).

5. Empirical Implications

5.1. Data

The data used for this study makes extensive use of the metadata assembled by the RePEc initiative and its various projects. 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.\(^\text{12}\)

In addition, we make use of the data made available by various projects that build on this RePEc data and enhance it in various ways. First, 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.\(^\text{13}\) Second, we get information about their advisors, students and alma mater, as recorded in the RePEc Genealogy project.\(^\text{14}\) Third, we gather in which mailing lists the papers have been disseminated through the NEP project.\(^\text{15}\) The latter have human editors determining to which field new working papers belong. Fourth, we make use of paper download data that is made available by the LogEc project.\(^\text{16}\) Fifth, we use citations to the papers and articles as extracted by the CitEc project.\(^\text{17}\) Sixth, we use journal impact factors, author and institution rankings from IDEAS.\(^\text{18}\) Finally, we make use of

\(^{12}\)See \url{http://repec.org/} for a general description of RePEc.
\(^{13}\)\url{https://authors.repec.org/}
\(^{14}\)\url{https://genealogy.repec.org/}
\(^{15}\)\url{https://nep.repec.org/}
\(^{16}\)\url{http://logec.repec.org/}
\(^{17}\)\url{http://citec.repec.org/}
\(^{18}\)\url{https://ideas.repec.org/top/}. For a detailed description of the factors and rankings, see Zimmermann
the “Ethnea” tool at the University of Illinois to establish the ethnicity of authors based on the first and last names.\textsuperscript{19}

The amount of data that is available for this project is overwhelming for the methods we need to adopt to estimate the model. For this reason, we apply a series of filters to reduce the sample size and to obtain records that are complete for our purposes:

1. We select papers which had a first working paper (pre-print) version for a given year, and do robustness exercises by repeating the computations for other individual years. As a benchmark year, we chose 2010 because it is old enough to give all authors the chance to have added the paper to their profile and for the paper to have been eventually published in a journal. But it is not too old to make sure we have a good-sized sample, as the coverage of RePEc becomes slimmer with older vintages.
2. We require all authors of the papers to be registered with RePEc.
3. We require that we can find in the RePEc Genealogy for all those authors where they studied and with which advisor(s).
4. We require that ethnicity could be determined for all authors.

In the end, we have a dataset for the years of 2010 to 2012 with 7,869 papers written by 3,345 distinct authors for which we have complete data. The numbers are similar for other years.\textsuperscript{20} In our empirical model, we use the weighted recursive discounted impact factors (IF) as the measure of a paper’s output. Since IF is measured from paper’s citations, we further drop 601 papers which do not have any citations up to January 2017 when retrieving from RePEc. Although it is natural to apply our methodology to this full sample with both single authored and co-authored papers, but computation is a problem remained to be solved. In the current implementation, we mainly focus on a subset of samples which are multi-authored. We call it "co-authored sample" and there are 2,389 papers and 1,826 authors involved. This co-authored sample would be mostly relevant due to the higher intensity of collaborations and potentially higher spillovers across authors. In addition, we extend this co-authored sample to accommodate the single-authored papers written by authors who are involved in the co-authored sample. This extended sample allows us to provide robustness check on results when authors’ efforts in their co-authored projects could be diluted by their efforts on own single-authored projects. To measure an author’s productivity, we use explanatory variables including author’s log life time citations (at the point of sample collection), years after his/her Ph.D. graduation, dummy variables for being a female, having the NBER affiliation, and graduating from the Ivy League. The summary statistics of variables that we use in our empirical model are provided in Table 1. A detailed description of variables can be found in Appendix E. Figure 3 shows the collaboration network among authors in the RePEc database.

\textsuperscript{19}http://abel.lis.illinois.edu/cgi-bin/ethnea/search.py
\textsuperscript{20}Summary statistics and estimation results covering the years 2007-2009 can be found in Appendix K.
Figure 3: The collaboration network among authors in the RePEc database considering only coauthored projects and dropping projects with zero impact factor. A node’s size and shade indicates its degree. The names of the five authors with the largest number of coauthors (degree) are indicated in the network.

Table 1: Summary statistics for the 2010-2012 sample.

<table>
<thead>
<tr>
<th></th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>S.D.</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Papers</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Citation recursive Impact Factor</td>
<td>0.0000</td>
<td>115.5851</td>
<td>6.5796</td>
<td>12.2021</td>
<td>3620</td>
</tr>
<tr>
<td>number of authors (in each paper)</td>
<td>1</td>
<td>5</td>
<td>1.8892</td>
<td>0.7108</td>
<td>3620</td>
</tr>
<tr>
<td><strong>Authors</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Log life-time citations</td>
<td>0</td>
<td>10.5516</td>
<td>5.4948</td>
<td>1.7118</td>
<td>1925</td>
</tr>
<tr>
<td>Decades after Ph.D. graduation</td>
<td>-0.6</td>
<td>9.9000</td>
<td>1.1113</td>
<td>0.9909</td>
<td>1925</td>
</tr>
<tr>
<td>Female</td>
<td>0</td>
<td>1</td>
<td>0.1345</td>
<td>0.3413</td>
<td>1925</td>
</tr>
<tr>
<td>NBER connection</td>
<td>0</td>
<td>1</td>
<td>0.1195</td>
<td>0.3244</td>
<td>1925</td>
</tr>
<tr>
<td>Ivy League connection</td>
<td>0</td>
<td>1</td>
<td>0.1553</td>
<td>0.3623</td>
<td>1925</td>
</tr>
<tr>
<td>Editor</td>
<td>0</td>
<td>1</td>
<td>0.0494</td>
<td>0.2167</td>
<td>1925</td>
</tr>
<tr>
<td>number of papers (for each author)</td>
<td>1</td>
<td>74</td>
<td>3.5527</td>
<td>3.8339</td>
<td>1925</td>
</tr>
</tbody>
</table>

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.
5.2. 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

\[
Y_s = \sum_{i=1}^{n} \delta_{i,s} e_{i,s} \alpha_i + \frac{\beta}{2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \delta_{i,s} \delta_{j,s} e_{i,s} e_{j,s} + \epsilon_s, \tag{19}
\]

where \( \delta_{i,s} \) is an indicator variable such that \( \delta_{i,s} = 1 \) if author \( i \) participates in paper \( s \) and \( \delta_{i,s} = 0 \) otherwise, \( e_{i,s} \) is the effort of author \( i \) put into paper \( s \), \( \alpha_i \) is a measure of the productivity of author \( i \) which may be modelled by \( \bm{x}_i \lambda \) with \( \bm{x}_i \) denotes a \((1 \times k)\)-vector of author characteristics, and \( \epsilon_s \) is a paper-specific random shock. The econometrician observes \( Y_s, \bm{x}_i, \) and \( \delta_{i,s}, \) but not \( e_{i,s} \) and \( \epsilon_s \).

The payoff function of author \( i \) is given by (cf. Equation (3))

\[
\pi_i = \sum_{s=1}^{p} \delta_{i,s} Y_s - \frac{1}{2} \left( \sum_{s=1}^{p} \delta_{i,s} e_{i,s}^2 + \phi \sum_{s=1}^{p} \sum_{r=1, r \neq s}^{p} \delta_{i,s} \delta_{i,r} e_{i,s} e_{i,r} \right). \tag{20}
\]

Substitution of (19) into (20) gives

\[
\pi_i = \sum_{s=1}^{p} \delta_{i,s} \left( \sum_{j=1}^{n} \delta_{j,s} e_{j,s} \xi_j + \frac{\beta}{2} \sum_{j=1, k \neq j}^{n} \delta_{j,s} \delta_{k,s} e_{j,s} e_{k,s} + \epsilon_s \right) - \frac{1}{2} \sum_{s=1}^{p} \delta_{i,s} \left( e_{i,s} + \phi \sum_{r=1, r \neq s}^{p} \delta_{i,r} e_{i,r} \right). \]

The FOC with respect to \( e_{i,s} \) is given by

\[
\frac{\partial \pi_i}{\partial e_{i,s}} = \delta_{i,s} \alpha_i + \beta \sum_{j=1, j \neq i}^{n} \delta_{j,s} e_{j,s} \xi_j - \delta_{i,s} e_{i,s} - \phi \sum_{r=1, r \neq s}^{p} \delta_{i,r} e_{i,r} = 0,
\]

which implies

\[
\delta_{i,s} e_{i,s} = \delta_{i,s} (\alpha_i + \beta \sum_{j=1, j \neq i}^{n} \delta_{j,s} e_{j,s} - \phi \sum_{r=1, r \neq s}^{p} \delta_{i,r} e_{i,r}), \tag{21}
\]
or

\[ \tilde{e}_{i,s} = \delta_{i,s}(\alpha_i + \beta \sum_{j=1, j \neq i}^{n} \tilde{e}_{j,s} - \phi \sum_{r=1, r \neq s}^{p} \tilde{e}_{i,r}), \tag{22} \]

where \( \tilde{e}_{i,s} = \delta_{i,s}e_{i,s} \). Note that (22) can be rewritten as

\[ (1 + \beta - \phi)\tilde{e}_{i,s} = \delta_{i,s}(\alpha_i + \beta \sum_{j=1}^{n} \tilde{e}_{j,s} - \phi \sum_{r=1}^{p} \tilde{e}_{i,r}). \tag{23} \]

Let \( \alpha = (\alpha_1, \ldots, \alpha_n)' \). Let \( \tilde{e}_s = (\tilde{e}_{1,s}, \ldots, \tilde{e}_{n,s})' \) and \( \tilde{e} = (\tilde{e}_1', \ldots, \tilde{e}_p')' \). Then, in vector form, (22) becomes

\[(1 + \beta - \phi)\tilde{e} = D(t_p \otimes \alpha + \beta(I_p \otimes \iota_n t_n')\tilde{e} - \phi(t_p t_p' \otimes I_n)\tilde{e}) \]

where \( D = \text{diag}_{s=1}^{p}\{\text{diag}_{i=1}^{n}\{\delta_{i,s}\}\} \) and \( \iota_n \) is an \( n \times 1 \) vector of ones. Hence, the equilibrium effort is given by

\[ \tilde{e}^* = [(1 + \beta - \phi)I_{np} - \beta D(I_p \otimes \iota_n t_n') + \phi D(t_p t_p' \otimes I_n)]^{-1}D(t_p \otimes \alpha). \tag{24} \]

Note that the equilibrium in Equation (24) is unique if

\[ S = [(1 + \beta - \phi)I_{np} - \beta D(I_p \otimes \iota_n t_n') + \phi D(t_p t_p' \otimes I_n)] \]

is nonsingular. Sufficient conditions for the nonsingularity of \( S \) can be found in Appendix F. Substitution of the equilibrium effort profile into (19) we obtain the predicted output of paper

\[ Y_{s}^* = \sum_{i=1}^{n} \tilde{e}_{i,s}^* \alpha_i + \frac{\beta}{2} \sum_{i=1}^{n} \sum_{j=1, j \neq i}^{n} \tilde{e}_{i,s}^* \tilde{e}_{j,s}^*. \tag{25} \]

The algorithm of solving equilibrium effort from Eq. (24) and predicted output from Eq. (25) allows us to estimate Eq. (19) by the nonlinear least squares (NLS) method or the maximum likelihood method when additional assumptions on \( \epsilon_s \) are given.

5.3. Matching Process

A problem with directly estimating Equation (25) is that the estimates of parameters may suffer from the biases due to endogeneity of \( D \). To address this endogeneity problem, we model the endogenous matching process of author \( i \) to paper \( s \) by

\[ \delta_{i,s} = 1\{\psi_{i,s} + u_{i,s} > 0\}, \]

where \( \psi_{i,s} \) denotes the matching quality between author \( i \) and paper \( s \) and \( u_{i,s} \) is a random component [cf. Chandrasekhar and Jackson, 2012]. We propose a specification to model \( \psi_{i,s} \), which is \( \psi_{i,s} = c_{i,s}'\gamma_1 + \gamma_2 \mu_i + \gamma_3 \kappa_s \). In this specification, \( c_{i,s} \) denotes a \( h \times 1 \) vector of dyad-specific regressors, which is to capture the homophily effect between the pair of author \( i \) and paper \( s \). In practice, from the RePEc data we only have one variable, which is the research field,
immediately available to reflect the similarity between authors and projects. In order to extend our control on the homophily effect, we regard the author in the project who ranks the highest in RePEc as the main author of the project, and thus construct the variable $c_{i,s}$ between the main author $s$ and author $i$ according to their similarities in gender, ethnicity, research fields, and whether they have an advisor-advisee relationship. The variable $\mu_i$ represents author $i$’s unobserved characteristic; and $\kappa_s$ represents paper $s$’s unobserved characteristic and they can reflect the feature of degree heterogeneity [cf. Graham, 2014, 2015]. Assuming $u_{i,s}$ is i.i.d. type-I extreme value distributed, we then obtain the following logistic regression model for the matching process:

$$\text{logit}(\delta_{i,s}) = c_{i,s}'\gamma_1 + \gamma_2\mu_i + \gamma_3\kappa_s.$$  \hspace{1cm} (26)

The key feature of the above endogenous matching equation is to introduce author and paper specific latent variables so that unobserved factors contributing to paper outputs can be controlled for. In other words, the production function of Eq. (19) should be modified into

$$Y_s = \sum_{i=1}^{n} \delta_{i,s} c_{i,s} \alpha_i + \frac{\beta}{2} \sum_{i=1}^{n} \sum_{j=1,j\neq i}^{n} \delta_{i,s} \delta_{j,s} c_{i,s} c_{j,s} + \eta \kappa_s + \nu_s, \hspace{1cm} \nu_s \sim N(0, \sigma^2_v), \hspace{1cm} (27)$$

where author’s productivity $\alpha_i$ can be modelled by $x_i \lambda + \rho \mu_i$. The joint probability function of $Y = (Y_1, \cdots, Y_p)$ and $D$ can be specified as

$$P(Y, D|x, c) = \int_{\mu} \int_{\kappa} P(Y|D, x, c, \mu, \kappa) P(D|c, \mu, \kappa) f(\mu) f(\kappa) d\mu d\kappa, \hspace{1cm} (28)$$

from which we can estimate the parameters $\theta = (\beta, \phi, \gamma, \lambda, \eta, \rho)$ with the author and paper latent variables, $\mu = (\mu_1, \cdots, \mu_n)$ and $\kappa = (\kappa_1, \cdots, \kappa_p)$.

### 5.4. Bayesian Estimation

Since the probability function of Eq. (28) involves a high dimensional integration of latent variables, it is not easy to apply the maximum likelihood method even when resorting to a simulation approach. As an alternative estimation method, the Bayesian MCMC approach can be more efficient for estimating latent variable models [cf. Zeger and Karim, 1991]. We divide the parameter $\theta$ and other unknown latent variables into blocks and assign the prior
distributions as follows:

\[ \mu_i \sim \mathcal{N}(0, 1), \quad i \in \mathcal{N} = \{1, \ldots, n\}, \]
\[ \kappa_s \sim \mathcal{N}(0, \sigma^2_k), \quad s \in \mathcal{P} = \{1, \ldots, p\}, \]
\[ \gamma = (\gamma_1, \gamma_2, \gamma_3) \sim \mathcal{N}_{k+2}(\gamma_0, \Gamma_0), \]
\[ \beta \sim \mathcal{N}(\beta_0, \Lambda_0), \]
\[ \phi \sim \mathcal{N}(\phi_0, \Phi_0), \]
\[ \xi = (\lambda, \rho) \sim \mathcal{N}_{k+1}(\xi_0, \Xi_0), \]
\[ \eta \sim \mathcal{N}(\eta_0, \Sigma_0); \quad \eta \geq 0, \]
\[ \sigma^2_k \sim \mathcal{IG}\left(\frac{\tau_0}{2}, \frac{\nu_0}{2}\right), \]
\[ \sigma^2_v \sim \mathcal{IG}\left(\frac{\tau_0}{2}, \frac{\nu_0}{2}\right). \]

We consider the normal and inverse gamma conjugate priors which are widely used in the Bayesian literature [Koop et al., 2007]. The hyper parameters are chosen to make the prior distribution relatively flat and cover a wide range of parameter space, i.e., we set \( \Lambda_0 = \Phi_0 = \Sigma_0 = 10; \Xi_0 = 10 \mathbf{I}_{k+1}; \Gamma_0 = 10 \mathbf{I}_k; \tau_0 = 2.2; \) and \( \nu_0 = 0.1. \)

The MCMC sampling consists of the following steps:

I. For \( i = 1, \ldots, n, \) update the latent variable \( \mu_i \) using the Metropolis-Hastings algorithm based on \( P(\mu_i|y, D). \)

II. Update \( \gamma \) using the Metropolis-Hastings algorithm based on \( P(\gamma|y, D). \)

III. For \( s = 1, \ldots, p, \) update the latent variable \( \kappa_s \) using the Metropolis-Hastings algorithm based on \( P(\kappa_s|y, D). \)

IV. Update \( \beta \) using the Metropolis-Hastings algorithm based on \( P(\beta|y, D). \)

V. Update \( \phi \) using the Metropolis-Hastings algorithm based on \( P(\phi|y, D). \)

VI. Update \( \xi \) using the Metropolis-Hastings algorithm based on \( P(\xi|y, D). \)

VII. Update \( \eta \) using the Metropolis-Hastings algorithm based on \( P(\eta|y, D). \)

VIII. Update \( \sigma^2_k \) and \( \sigma^2_v \) using conjugate inverse gamma conditional posterior distributions.

5.5. Estimation Results

We report estimation results for both cases of homogeneous and heterogeneous spillovers in Table 2. In each case, the first column, i.e., Model (I), shows the results where we have assumed that the collaboration network is exogenously given, and the estimation procedure is solely based on the production function outlined in Section 5.2. The second column, i.e., Model (II), allows the collaboration network to be formed endogenously, and is based on the joint estimation of the production function and the matching process in Section 5.3.
First of all, in case of homogeneous spillover, we find in Model (I) that the spillover effect of efforts between co-authors, measured by $\beta$, does not have the expected positive sign. In addition, the congestion effect between projects, measured by $\phi$, is insignificant. When comparing to the results in Model (II), where the estimates of $\beta$ and $\phi$ are both significant and having the expected signs, we conclude that the estimates of $\beta$ and $\phi$ in Model (I) are downward biased due to the problem of endogenous matching between authors and projects. To show why biases are downward, we provide a heuristic explanation in Appendix G by using the estimates from Model (I) and Model (II) to simulate author abilities, efforts, and predicted paper outputs. We show that if the estimates of $\beta$ and $\phi$ in Model (I) become higher than what they currently are, the predicted paper outputs will further deviate from the true ones. In addition, we conduct a Monte Carlo simulation study to investigate the performance of our estimation method. As shown by the simulation results in Appendix H, we also see the same downward biases on the estimates of $\beta$ and $\phi$ when the collaboration network was treated exogenous mistakenly.

Speaking to the effect of author characteristics in paper output, we also find correcting the problem of endogenous matching would change some estimates from Model (I) to Model (II). Based on the results in Model (II), the coefficient of the number of lifetime citations (0.7293) is a positive and significantly predictor [cf. e.g. Ductor, 2014]. We follow Rauber and Ursprung [2008] and Krapf et al. [2017] to capture career experience (measured by decades after Ph.D. graduation) in a polynomial of order five. The result shows that the coefficient of the first order (-1.4071) is significantly negative, while the rest higher orders are insignificant. Female dummy (0.2907) is positively affecting output [cf. Ductor et al., 2017; Krapf et al., 2017]. Being affiliated with the NBER (1.1850) positively and significantly impacts research output. Having attended an Ivy League university (0.4761) also positively affects output. The editor dummy has a negative but insignificant effect on output. The author-specific latent variable (4.7197) is found positively and significantly affecting author’s productivity. There is no significant effect from paper’s latent variable on paper’s output.

For the matching between authors and projects, we find that having the same ethnicity, having the same affiliation, being past co-authors, and sharing common co-authors with the lead researcher of a publication make matching more likely [cf. Freeman and Huang, 2015]. Similarities in the NEP fields also positively and significantly affect matchings [Ductor, 2014]. Being in a Ph.D. advisor–advisee relationship also largely contributes to matchings. Finally, author’s latent variable shows a positively significant effect on the author-project matching.

In case of the heterogeneous spillover, we construct the Jaffe proximity measures of research fields (NEP) between each pair of authors and then incorporate the proximity measure into the production function of Eq. (1). Appendix J shows how to modify the computation of equilibrium efforts from this heterogeneous specification. In the second case of Table 2 we again find that when omitting the endogenous matching of authors and papers, the estimate of $\beta$ and $\phi$ are downward biased. Also, none of author characteristics show significant coefficients. After coping with the endogenous matching in the full model, the estimate of $\beta$ resumes significant and shows a slightly larger value compared to the homogeneous spillover case; meanwhile, the estimate of
\( \phi \) also becomes significant but has a slightly smaller value compared to the homogeneous case.

6. Rankings for Individuals and Departments

With our estimates from the previous section (cf. Table 2) we are now able to perform various counterfactual studies. We first investigate the reduction in total output upon the removal of individual authors or entire departments from the network (cf. Section 4.1).

The ranking of individual authors and departments can be found in Tables 3 and 4, respectively. The key author turns out to be Gianmarco Ottaviano from the London School of Economics. Our results suggest that without this author total output would be about 1.4% lower. The second and third highest ranked authors are Carmen Reinhart from Harvard University and Dirk Bergemann from Yale University. Their impact on research output is similarly high. The London School of Economics and Harvard University are also among the top three ranked institutions in Table 4. From the ranking of authors we observe that highly ranked authors tend to have a larger number of papers, a larger number of citations or a higher RePEc rank, but these indicators do not yield the same ranking that we obtain based on our model and the data. However, this discrepancy is not surprising, as other rankings are typically not derived from microeconomic foundations, and do not take into account spillover effects generated in scientific knowledge production networks.
Table 2: Estimation results for the 2010-2012 sample

<table>
<thead>
<tr>
<th>Output</th>
<th>Homogeneous Spillover</th>
<th>Heterogeneous Spillover</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (I)</td>
<td>Model (II)</td>
</tr>
<tr>
<td></td>
<td>Model (I)</td>
<td>Model (II)</td>
</tr>
<tr>
<td>β</td>
<td>-0.0725**</td>
<td>-0.0420</td>
</tr>
<tr>
<td></td>
<td>(0.0358)</td>
<td>(0.0069)</td>
</tr>
<tr>
<td></td>
<td>0.0666***</td>
<td>0.0785**</td>
</tr>
<tr>
<td></td>
<td>(0.0171)</td>
<td>(0.0031)</td>
</tr>
<tr>
<td>ϕ</td>
<td>0.0092</td>
<td>0.0056</td>
</tr>
<tr>
<td></td>
<td>(0.0057)</td>
<td>(0.0059)</td>
</tr>
<tr>
<td></td>
<td>0.0634***</td>
<td>0.0526**</td>
</tr>
<tr>
<td></td>
<td>(0.0107)</td>
<td>(0.0105)</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.4599***</td>
<td>-0.9804</td>
</tr>
<tr>
<td></td>
<td>(0.5092)</td>
<td>(0.8142)</td>
</tr>
<tr>
<td></td>
<td>-2.6928***</td>
<td>-2.6607**</td>
</tr>
<tr>
<td></td>
<td>(0.2450)</td>
<td>(0.2433)</td>
</tr>
<tr>
<td>Log life-time citations</td>
<td>0.0275***</td>
<td>0.0785</td>
</tr>
<tr>
<td></td>
<td>(0.1800)</td>
<td>(0.0548)</td>
</tr>
<tr>
<td></td>
<td>0.7293***</td>
<td>0.7319**</td>
</tr>
<tr>
<td></td>
<td>(0.4130)</td>
<td>(0.0516)</td>
</tr>
<tr>
<td>Decades after graduation</td>
<td>-0.3558</td>
<td>-0.2336</td>
</tr>
<tr>
<td></td>
<td>(0.5622)</td>
<td>(0.4810)</td>
</tr>
<tr>
<td></td>
<td>-1.4071***</td>
<td>-1.0809**</td>
</tr>
<tr>
<td></td>
<td>(0.4452)</td>
<td>(0.3822)</td>
</tr>
<tr>
<td></td>
<td>(Decades after graduation)^2</td>
<td>-0.8704</td>
</tr>
<tr>
<td></td>
<td>(Decades after graduation)^3</td>
<td>0.4845</td>
</tr>
<tr>
<td></td>
<td>(Decades after graduation)^4</td>
<td>-0.0893</td>
</tr>
<tr>
<td></td>
<td>Female</td>
<td>0.3512</td>
</tr>
<tr>
<td></td>
<td>0.2907*</td>
<td>0.2466</td>
</tr>
<tr>
<td></td>
<td>(0.2122)</td>
<td>(0.2507)</td>
</tr>
<tr>
<td></td>
<td>0.3684***</td>
<td>0.4303**</td>
</tr>
<tr>
<td></td>
<td>(0.1695)</td>
<td>(0.1691)</td>
</tr>
<tr>
<td>NBER connection</td>
<td>0.7750***</td>
<td>0.4988</td>
</tr>
<tr>
<td></td>
<td>(0.2498)</td>
<td>(0.4048)</td>
</tr>
<tr>
<td></td>
<td>1.1850***</td>
<td>1.1604**</td>
</tr>
<tr>
<td></td>
<td>(0.1380)</td>
<td>(0.1480)</td>
</tr>
<tr>
<td>Ivy League connection</td>
<td>0.4937***</td>
<td>0.3272</td>
</tr>
<tr>
<td></td>
<td>(0.1825)</td>
<td>(0.2749)</td>
</tr>
<tr>
<td></td>
<td>0.4761***</td>
<td>0.5184**</td>
</tr>
<tr>
<td></td>
<td>(0.1203)</td>
<td>(0.1260)</td>
</tr>
<tr>
<td>Editor</td>
<td>-0.1259</td>
<td>-0.0781</td>
</tr>
<tr>
<td></td>
<td>(0.2131)</td>
<td>(0.1794)</td>
</tr>
<tr>
<td></td>
<td>0.2271</td>
<td>0.1479</td>
</tr>
<tr>
<td>ρ</td>
<td>4.7197*</td>
<td>4.7740**</td>
</tr>
<tr>
<td>η</td>
<td>0.0369</td>
<td>0.1948</td>
</tr>
<tr>
<td></td>
<td>(0.7930)</td>
<td>(0.8872)</td>
</tr>
<tr>
<td>σ²</td>
<td>108.6973***</td>
<td>116.9718***</td>
</tr>
<tr>
<td></td>
<td>(2.9484)</td>
<td>(2.7905)</td>
</tr>
<tr>
<td></td>
<td>83.5926***</td>
<td>89.6188***</td>
</tr>
<tr>
<td></td>
<td>(2.0955)</td>
<td>(2.1760)</td>
</tr>
<tr>
<td></td>
<td>(2.7905)</td>
<td>(2.1760)</td>
</tr>
</tbody>
</table>

Matching constant –9.7215*** –9.7766***
Same NEP –0.7446*** –0.7233***
Ethnicity –0.9138*** –0.9444***
Affiliation –3.8666*** –3.9564***
Gender –0.0226 –0.0282
Advisor-advisee –3.8669*** –3.9967***
Past coauthors 5.9871*** 6.0566***
Share common co-authors –3.2517*** –3.3309***
Author effect –4.5616*** –5.4911***
Project effect –0.0899 –0.1313

Sample size 3620 3620

Note: Model (1): assume exogenous matching between authors and papers. Model (2): assume endogenous matching by Equation (26). The asterisks ‘***’ indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.
Table 3: Ranking of the top-twenty five researchers from the 2010-2012 sample.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Name</th>
<th>Projects</th>
<th>Citations</th>
<th>RePEc Rank$^a$</th>
<th>Output Loss$^b$</th>
<th>Institute</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Ottaviano, Gianmarco I.P.</td>
<td>14</td>
<td>4384</td>
<td>198</td>
<td>-1.359%</td>
<td>Economics Department, London School of Economics (LSE)</td>
</tr>
<tr>
<td>2</td>
<td>Reinhart, Carmen M.</td>
<td>8</td>
<td>17004</td>
<td>21</td>
<td>-1.309%</td>
<td>Kennedy School of Government, Harvard University</td>
</tr>
<tr>
<td>3</td>
<td>Bergemann, Dirk</td>
<td>32</td>
<td>1007</td>
<td>968</td>
<td>-1.198%</td>
<td>Economics Department, Yale University</td>
</tr>
<tr>
<td>4</td>
<td>Pischke, Jorn-Steffen</td>
<td>5</td>
<td>1194</td>
<td>516</td>
<td>-1.045%</td>
<td>Centre for Economic Performance (CEP), London School of Economics (LSE)</td>
</tr>
<tr>
<td>5</td>
<td>Mayer, Thierry</td>
<td>13</td>
<td>2896</td>
<td>482</td>
<td>-1.022%</td>
<td>Department of Economics, Sciences Economiques, Sciences Po</td>
</tr>
<tr>
<td>6</td>
<td>Perri, Fabrizio</td>
<td>11</td>
<td>3506</td>
<td>752</td>
<td>-0.936%</td>
<td>Research Department, Federal Reserve Bank of Minneapolis</td>
</tr>
<tr>
<td>7</td>
<td>Giuliano, Paola</td>
<td>7</td>
<td>2483</td>
<td>1515</td>
<td>-0.896%</td>
<td>University of California-Los Angeles (UCLA)</td>
</tr>
<tr>
<td>8</td>
<td>van Reenen, John M.</td>
<td>17</td>
<td>1467</td>
<td>78</td>
<td>-0.854%</td>
<td>Centre for Economic Performance (CEP), London School of Economics (LSE)</td>
</tr>
<tr>
<td>9</td>
<td>Nunn, Nathan</td>
<td>10</td>
<td>1914</td>
<td>635</td>
<td>-0.852%</td>
<td>Department of Economics, Harvard University</td>
</tr>
<tr>
<td>10</td>
<td>Black, Sandra E.</td>
<td>5</td>
<td>382</td>
<td>645</td>
<td>-0.827%</td>
<td>Department of Economics, University of Texas-Austin</td>
</tr>
<tr>
<td>11</td>
<td>Angrist, Joshua D.</td>
<td>6</td>
<td>6168</td>
<td>51</td>
<td>-0.818%</td>
<td>Economics Department, Massachusetts Institute of Technology (MIT)</td>
</tr>
<tr>
<td>12</td>
<td>Michalopoulos, Stelios</td>
<td>11</td>
<td>8099</td>
<td>2141</td>
<td>-0.817%</td>
<td>Economics Department, Brown University</td>
</tr>
<tr>
<td>13</td>
<td>Bloom, Nicholas</td>
<td>12</td>
<td>1091</td>
<td>210</td>
<td>-0.746%</td>
<td>Department of Economics, Stanford University</td>
</tr>
<tr>
<td>14</td>
<td>Levchenko, Andrei A.</td>
<td>9</td>
<td>2368</td>
<td>1230</td>
<td>-0.740%</td>
<td>Economics Department, University of Michigan</td>
</tr>
<tr>
<td>15</td>
<td>Huizinga, Harry P.</td>
<td>12</td>
<td>3857</td>
<td>729</td>
<td>-0.730%</td>
<td>School of Economics and Management, Universiteit van Tilburg</td>
</tr>
<tr>
<td>16</td>
<td>Spolaore, Enrico</td>
<td>6</td>
<td>1257</td>
<td>1303</td>
<td>-0.712%</td>
<td>Department of Economics, Tufts University</td>
</tr>
<tr>
<td>17</td>
<td>Sufi, Amir</td>
<td>7</td>
<td>3969</td>
<td>1309</td>
<td>-0.709%</td>
<td>Booth School of Business, University of Chicago</td>
</tr>
<tr>
<td>18</td>
<td>Saez, Emmanuel</td>
<td>8</td>
<td>2451</td>
<td>306</td>
<td>-0.685%</td>
<td>Department of Economics, University of California-Berkeley</td>
</tr>
<tr>
<td>19</td>
<td>Basu, Susanto</td>
<td>3</td>
<td>1105</td>
<td>682</td>
<td>-0.628%</td>
<td>Department of Economics, Boston College</td>
</tr>
<tr>
<td>20</td>
<td>Dohmen, Thomas J.</td>
<td>5</td>
<td>1451</td>
<td>1081</td>
<td>-0.561%</td>
<td>Wirtschaftswissenschaftlicher Fachbereich, University of Bonn</td>
</tr>
<tr>
<td>21</td>
<td>Pathak, Parag</td>
<td>3</td>
<td>1230</td>
<td>1063</td>
<td>-0.555%</td>
<td>National Bureau of Economic Research (NBER)</td>
</tr>
<tr>
<td>22</td>
<td>Bandiera, Oriana</td>
<td>6</td>
<td>2094</td>
<td>1028</td>
<td>-0.544%</td>
<td>Economics Department, London School of Economics (LSE)</td>
</tr>
<tr>
<td>23</td>
<td>Adam, Klaus</td>
<td>10</td>
<td>813</td>
<td>1985</td>
<td>-0.541%</td>
<td>Center for Financial Studies</td>
</tr>
<tr>
<td>24</td>
<td>Rogoff, Kenneth</td>
<td>8</td>
<td>7565</td>
<td>8</td>
<td>-0.537%</td>
<td>Department of Economics, Harvard University</td>
</tr>
<tr>
<td>25</td>
<td>Quadrini, Vincenzo</td>
<td>8</td>
<td>1466</td>
<td>1299</td>
<td>-0.503%</td>
<td>Marshall School of Business, University of Southern California</td>
</tr>
</tbody>
</table>

$^a$ The RePEc ranking is based on an aggregate of rankings by different criteria. See Zimmermann [2013] for more information.

$^b$ 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 2. See also Equation (4.1) in Section 10.
Table 4: Ranking of the top-ten departments from the 2010-2012 sample.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Department/School</th>
<th>Size</th>
<th>RePEc Ranka</th>
<th>Output Lossb</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Kennedy School of Government, Harvard University</td>
<td>9</td>
<td>16</td>
<td>-2.667%</td>
</tr>
<tr>
<td>2</td>
<td>Economics Department, London School of Economics</td>
<td>9</td>
<td>33</td>
<td>-2.620%</td>
</tr>
<tr>
<td>3</td>
<td>Booth School of Business, University of Chicago</td>
<td>11</td>
<td>8</td>
<td>-2.375%</td>
</tr>
<tr>
<td>4</td>
<td>Department of Economics, Harvard University</td>
<td>20</td>
<td>1</td>
<td>-2.318%</td>
</tr>
<tr>
<td>5</td>
<td>Centre for Economic Performance, London School of Economics</td>
<td>6</td>
<td>75</td>
<td>-2.188%</td>
</tr>
<tr>
<td>6</td>
<td>Department of Economics, Princeton University</td>
<td>11</td>
<td>7</td>
<td>-2.125%</td>
</tr>
<tr>
<td>7</td>
<td>Department of Economics, Sciences Economiques, Sciences Po</td>
<td>5</td>
<td>115</td>
<td>-2.063%</td>
</tr>
<tr>
<td>8</td>
<td>Economics Department, University of Michigan</td>
<td>12</td>
<td>30</td>
<td>-1.776%</td>
</tr>
<tr>
<td>9</td>
<td>Economics Department, Brown University</td>
<td>12</td>
<td>21</td>
<td>-1.641%</td>
</tr>
<tr>
<td>10</td>
<td>National Bureau of Economic Research</td>
<td>17</td>
<td>2</td>
<td>-1.615%</td>
</tr>
</tbody>
</table>

a The RePEc ranking is based on an aggregate of rankings by different criteria. See Zimmermann [2013] for more information.
b The output loss for department $D$ is computed as $\sum_{s=1}^{p} Y_s(G) - \sum_{s=1}^{p} Y_s(G\backslash D)$ with the parameter estimates from Table 2. See also Equation (11) in Section 10.

7. Research Funding for Individuals and Departments

In this section we compare our optimal research funding scheme $z^*$ of Equation (18) in Section 4.2 using the parameter estimates from Section 5.5 with funding programs being implemented in the real world [cf. e.g. De Frajy, 2016; Stephan, 2012]. For this purpose 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. from 1976 to 2016 [cf. Drutman, 2012].21,22

The economist receiving the largest amount of funds from the NSF is Frank Stafford from the University of Michigan with total funds amounting to 33,471,414 U.S. dollars. He manages the Panel Study of Income Dynamics (PSID) of U.S. families, which was among the NSF “Top Sixty” overall funded programs in 2010. The average funding amount from the NSF is 436,201 U.S. dollars. Figure 5 shows the total awards from the NSF to economists together with the highly skewed distribution of awards over the years ranging from 2000 to 2010. At the level of organizations and departments, 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. The average funding across organizations from the NSF is 2,831,612 U.S. dollars.

Figure 6 shows the optimal relative funding solving the planner’s problem stated in Equation (18) with the estimated parameters from Table 2 sorted across authors. The figure illustrates that the optimal funding policy is highly skewed and tends to concentrate funds towards the most productive authors. The large heterogeneity of funds across researchers is reminiscent of the heterogeneous distribution of NSF awards that we have observed in Figure 5.

21 See https://www.nsf.gov/awardsearch/.
22 The data coverage before 1976 is incomplete, and we thus discarded years prior to 1976.
Figure 5: (Left panel) The total awards from the NSF to economists and (right panel) the distribution of awards over the years 2000 to 2010.

Figure 6: The optimal relative funding sorted across authors.
However, these similarities do not carry over to the level of individual authors or departments. Table 5 shows the optimal research funding per individual together with the awards these authors actually received from the NSF relative to the total awards provided by the NSF. We find that the rankings of our optimal funding policy and the one by the NSF differ. The author with the highest funds according to our optimal policy is Dirk Bergemann from Yale University (with 5.29% of the total funds) followed by Nicholas Bloom from Stanford University (with 4.62% of the total funds). However, the second ranked researcher received more than three times as much funding from the NSF as the first ranked researcher. The difference between the optimal funding policy that we obtain and the one implemented by the NSF is, however, not surprising, as current research funding instruments typically do not take into account the spillover effects generated in scientific knowledge production networks. Moreover, similar to the author ranking of Table 3, we find that the ranking based on the optimal funding policy does not coincide with the ranking based on citations or the RePEc ranking. As discussed above, the reason being that the optimal funding policy that we analyze explicitly takes spillovers in the coauthorship network into account.
Table 5: Ranking of the optimal research funding for the top-twenty five researchers for the 2010-2012 sample.\textsuperscript{a}

<table>
<thead>
<tr>
<th>Name</th>
<th>Projects</th>
<th>Citations</th>
<th>RePEc Rank\textsuperscript{b}</th>
<th>Institution</th>
<th>NSF [%]</th>
<th>Funding [%]\textsuperscript{c}</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dirk Bergemann</td>
<td>36</td>
<td>1018</td>
<td>951</td>
<td>Economics Department, Yale University</td>
<td>0.9060</td>
<td>5.2928</td>
<td>1</td>
</tr>
<tr>
<td>Nicholas Bloom</td>
<td>13</td>
<td>4202</td>
<td>188</td>
<td>Department of Economics, Stanford University</td>
<td>2.9818</td>
<td>4.6210</td>
<td>2</td>
</tr>
<tr>
<td>Stephen Morris</td>
<td>31</td>
<td>3414</td>
<td>284</td>
<td>Department of Economics, Princeton University</td>
<td>2.1517</td>
<td>4.1196</td>
<td>3</td>
</tr>
<tr>
<td>John List</td>
<td>29</td>
<td>7741</td>
<td>27</td>
<td>Department of Economics, University of Chicago</td>
<td>0.1333</td>
<td>4.0159</td>
<td>4</td>
</tr>
<tr>
<td>Fabrizio Perri</td>
<td>11</td>
<td>1909</td>
<td>738</td>
<td>Research Department, Federal Reserve Bank of Minneapolis</td>
<td>0.4135</td>
<td>3.9260</td>
<td>5</td>
</tr>
<tr>
<td>Oded Galor</td>
<td>17</td>
<td>7663</td>
<td>84</td>
<td>Economics Department, Brown University</td>
<td>0.8217</td>
<td>3.8178</td>
<td>6</td>
</tr>
<tr>
<td>Craig Burnside</td>
<td>10</td>
<td>2700</td>
<td>578</td>
<td>Department of Economics, Duke University</td>
<td>0.4257</td>
<td>3.7843</td>
<td>7</td>
</tr>
<tr>
<td>Sergio Rebele</td>
<td>9</td>
<td>8043</td>
<td>127</td>
<td>Centre for Economic Policy Research (CEPR)</td>
<td>0.8901</td>
<td>3.6637</td>
<td>8</td>
</tr>
<tr>
<td>Emmanuel Saez</td>
<td>14</td>
<td>3930</td>
<td>314</td>
<td>Department of Economics, University of California-Berkeley</td>
<td>2.7857</td>
<td>3.4432</td>
<td>9</td>
</tr>
<tr>
<td>Martin Eichenbaum</td>
<td>7</td>
<td>10252</td>
<td>68</td>
<td>Department of Economics, Northwestern University</td>
<td>0.4995</td>
<td>3.2291</td>
<td>10</td>
</tr>
<tr>
<td>Vincenzo Quadrini</td>
<td>8</td>
<td>1460</td>
<td>1359</td>
<td>Department of Economics, University of Southern California</td>
<td>1.8360</td>
<td>3.0780</td>
<td>11</td>
</tr>
<tr>
<td>Joshua Angrist</td>
<td>6</td>
<td>8230</td>
<td>53</td>
<td>Economics Department, Massachusetts Institute of Technology (MIT)</td>
<td>2.5971</td>
<td>3.0176</td>
<td>12</td>
</tr>
<tr>
<td>Andrei Levchenko</td>
<td>12</td>
<td>1081</td>
<td>1120</td>
<td>Economics Department, University of Michigan</td>
<td>0.5309</td>
<td>2.9679</td>
<td>13</td>
</tr>
<tr>
<td>Sandra Black</td>
<td>5</td>
<td>2813</td>
<td>563</td>
<td>Department of Economics, University of Texas-Austin</td>
<td>0.5879</td>
<td>2.7213</td>
<td>14</td>
</tr>
<tr>
<td>Chad Syverson</td>
<td>6</td>
<td>1656</td>
<td>574</td>
<td>Booth School of Business, University of Chicago</td>
<td>0.9976</td>
<td>2.4672</td>
<td>15</td>
</tr>
<tr>
<td>Yuriy Gorodnichenko</td>
<td>8</td>
<td>1940</td>
<td>495</td>
<td>Department of Economics, University of California-Berkeley</td>
<td>0.8387</td>
<td>2.2451</td>
<td>16</td>
</tr>
<tr>
<td>Marc Melitz</td>
<td>10</td>
<td>6763</td>
<td>145</td>
<td>Department of Economics, Harvard University</td>
<td>0.5976</td>
<td>1.8869</td>
<td>17</td>
</tr>
<tr>
<td>Kenneth Rogoff</td>
<td>9</td>
<td>21001</td>
<td>8</td>
<td>Department of Economics, Harvard University</td>
<td>1.1896</td>
<td>1.7957</td>
<td>18</td>
</tr>
<tr>
<td>Erik Snowberg</td>
<td>9</td>
<td>153</td>
<td>9054</td>
<td>Division of Social Sciences, California Institute of Technology</td>
<td>0.5109</td>
<td>1.4143</td>
<td>19</td>
</tr>
<tr>
<td>David Autor</td>
<td>6</td>
<td>5006</td>
<td>139</td>
<td>Economics Department, Massachusetts Institute of Technology (MIT)</td>
<td>2.0530</td>
<td>1.3558</td>
<td>20</td>
</tr>
<tr>
<td>Matthias Doepke</td>
<td>6</td>
<td>1273</td>
<td>935</td>
<td>Department of Economics, Northwestern University</td>
<td>1.4061</td>
<td>1.3177</td>
<td>21</td>
</tr>
<tr>
<td>Thomas Holmes</td>
<td>8</td>
<td>1102</td>
<td>1033</td>
<td>Department of Economics, University of Minnesota</td>
<td>0.9731</td>
<td>1.2300</td>
<td>22</td>
</tr>
<tr>
<td>Gordon Hanson</td>
<td>5</td>
<td>5346</td>
<td>194</td>
<td>Department of Economics, University of California-San Diego (UCSD)</td>
<td>0.8390</td>
<td>1.1805</td>
<td>23</td>
</tr>
<tr>
<td>Dean Karlan</td>
<td>8</td>
<td>1968</td>
<td>742</td>
<td>Economics Department, Yale University</td>
<td>2.8064</td>
<td>1.1255</td>
<td>24</td>
</tr>
<tr>
<td>George-Marios Angeles</td>
<td>8</td>
<td>1678</td>
<td>759</td>
<td>Economics Department, Massachusetts Institute of Technology (MIT)</td>
<td>0.5126</td>
<td>1.0483</td>
<td>25</td>
</tr>
</tbody>
</table>

\textsuperscript{a} We only consider 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.

\textsuperscript{b} The RePEc ranking is based on an aggregate of rankings by different criteria. See Zimmermann [2013] for more information.

\textsuperscript{c} The total cost of funds, $\sum_{s=1}^{p} \delta_{i,s} \bar{Y}_{s}(G, \bar{z})$, of researcher $i$ with the optimal research funding scheme $\bar{z}$ of Equation (18) in Section 4.2 with the parameter estimates from Table 2.
Figure 7: 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. The data have been log transformed to account for the heterogeneity across observations.

Figure 7 shows the correlations of the authors’ degrees, lifetime citations, total NSF awards and the optimal funding policy. We observe that the optimal funding policy is significantly positively correlated with the degree (number of coauthors). In contrast, the NSF awards are positively but not significantly correlated with the degree or the optimal funding policy. This highlights the importance of the collaboration network in determining the optimal funding policy, while it does not seem to have an effect on the allocation of NSF awards.

A similar ranking as in Table 5, but at the departmental level, can be found in Table 6. We find that the Department of Economics at Yale University receives the largest amount of funding, followed by the University of California at Berkeley. Similar to the ranking of individual authors in Table 5, we observe that the actual funding provided by the NSF does not coincide with the optimal funding policy that we obtain, which explicitly considers spillover effects between the authors of different departments.

8. Conclusion

In this paper we have analyzed the equilibrium efforts of authors involved in multiple, possibly overlapping projects. We show that, given an allocation of researchers to different projects, the Nash equilibrium can be completely characterized. We then 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
Table 6: Ranking of optimal research funding for the top-ten departments for the 2010-2012 sample.a

<table>
<thead>
<tr>
<th>Institution</th>
<th>Size</th>
<th>NSF [%]</th>
<th>Funding [%]</th>
<th>Rank</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yale University</td>
<td>22</td>
<td>2.2730</td>
<td>9.7595</td>
<td>1</td>
</tr>
<tr>
<td>University of California-Berkeley</td>
<td>24</td>
<td>2.0600</td>
<td>7.6783</td>
<td>2</td>
</tr>
<tr>
<td>University of Chicago</td>
<td>30</td>
<td>2.0529</td>
<td>7.1327</td>
<td>3</td>
</tr>
<tr>
<td>Massachusetts Institute of Technology</td>
<td>18</td>
<td>1.4370</td>
<td>7.0252</td>
<td>4</td>
</tr>
<tr>
<td>Harvard University</td>
<td>46</td>
<td>1.6506</td>
<td>6.2985</td>
<td>5</td>
</tr>
<tr>
<td>Northwestern University</td>
<td>18</td>
<td>3.7579</td>
<td>5.8245</td>
<td>6</td>
</tr>
<tr>
<td>Princeton University</td>
<td>14</td>
<td>2.5822</td>
<td>5.4320</td>
<td>7</td>
</tr>
<tr>
<td>Stanford University</td>
<td>21</td>
<td>3.5354</td>
<td>4.6851</td>
<td>8</td>
</tr>
<tr>
<td>University of Michigan</td>
<td>22</td>
<td>0.5603</td>
<td>4.3060</td>
<td>9</td>
</tr>
<tr>
<td>Brown University</td>
<td>15</td>
<td>0.7620</td>
<td>3.9831</td>
<td>10</td>
</tr>
</tbody>
</table>

a We only consider 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, \( \sum_{i \in D} \sum_{s=1}^{n_s} \delta_{is} z^* Y_s(G, z) \), for each department \( D \) and researchers \( i \in D \) with the optimal research funding scheme \( z^* \) of Equation (18) in Section 4.2 with the parameter estimates from Table 2.

...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.

Our analysis can be extended along several directions. First, we can allow the returns of an author from participating in a project to be split equally among the participants of the project similar to the models studied in Jackson and Wolinsky [1996]; Kandel and Lazear [1992]. Second, instead of a convex cost, we can introduce a time constraint as in Baumann [2014] and Salonen [2016].23 Third, we can compare our optimal research funding policy with the ones implemented in practice not only in the U.S. but also in other countries. 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.

References


23 These extensions and the relation to the current setup are summarized in Table 7 in Appendix B.


Appendix

A. Additional Examples

In the following we provide two more examples illustrating the Nash equilibrium characterization of Propositions 1 and 2 for specific networks.

Example 2. Consider a network with 2 projects and 4 agents, where in the first project agents 1, 2 and 3 are collaborating while in the second project agents 2 and 4 are collaborating. An illustration can be found in Figure A.1. The payoffs of the agents are given by

\[
\begin{align*}
\pi_1 &= e_{11} \left( \alpha_1 + \frac{\beta}{2} (e_{21} + e_{31}) \right) + e_{21} \left( \alpha_2 + \frac{\beta}{2} (e_{11} + e_{31}) \right) + e_{31} \left( \alpha_3 + \frac{\beta}{2} (e_{11} + e_{21}) \right) - \frac{\gamma}{2} e_{11}^2 \\
\pi_2 &= e_{21} \left( \alpha_2 + \frac{\beta}{2} (e_{11} + e_{31}) \right) + e_{11} \left( \alpha_1 + \frac{\beta}{2} (e_{21} + e_{31}) \right) + e_{31} \left( \alpha_3 + \frac{\beta}{2} (e_{11} + e_{21}) \right) + e_{22} \left( \alpha_2 + \frac{\beta}{2} e_{42} \right) \\
&\quad + e_{42} \left( \alpha_4 + \frac{\beta}{2} e_{22} \right) - \frac{\gamma}{2} e_{21}^2 - \frac{\gamma}{2} e_{22}^2 - \rho e_{21} e_{22} \\
\pi_3 &= e_{31} \left( \alpha_3 + \frac{\beta}{2} (e_{11} + e_{21}) \right) + e_{21} \left( \alpha_1 + \frac{\beta}{2} (e_{21} + e_{31}) \right) + e_{21} \left( \alpha_2 + \frac{\beta}{2} (e_{11} + e_{31}) \right) - \frac{\gamma}{2} e_{31}^2 \\
\pi_4 &= e_{42} \left( \alpha_4 + \frac{\beta}{2} e_{22} \right) + e_{22} \left( \alpha_2 + \frac{\beta}{2} e_{42} \right) - \frac{\gamma}{2} e_{42}^2.
\end{align*}
\]

The first order conditions are given by

\[
\begin{align*}
\frac{\partial \pi_1}{\partial e_{11}} &= \alpha_1 + (e_{21} + e_{31}) \beta - e_{11} \gamma = 0 \\
\frac{\partial \pi_2}{\partial e_{21}} &= \alpha_2 + (e_{11} + e_{31}) \beta - e_{21} \gamma - e_{22} \rho = 0 \\
\frac{\partial \pi_2}{\partial e_{22}} &= \alpha_2 + e_{22} \beta - e_{22} \gamma - e_{21} \rho = 0 \\
\frac{\partial \pi_3}{\partial e_{31}} &= \alpha_3 + (e_{11} + e_{21}) \beta - e_{31} \gamma = 0 \\
\frac{\partial \pi_4}{\partial e_{42}} &= \alpha_4 + e_{22} \beta - e_{42} \gamma = 0.
\end{align*}
\]

Solving this system of equations directly yields

\[
\begin{align*}
e_{11} &= \frac{-(\alpha_2 + \alpha_3) \beta + \alpha_1 (\beta - \gamma)) (\beta - \gamma)(\beta + \gamma)^2 - \beta (\beta + \gamma) (\alpha_4 \beta + \alpha_2 \gamma) \rho - \gamma (\alpha_3 \beta + \alpha_1 \gamma) \rho^2}{(\beta - \gamma)(\beta + \gamma) (2 \beta - \gamma)(\beta + \gamma)^2 + \gamma \rho^2}, \\
e_{21} &= \frac{- (\beta + \gamma) ((\alpha_1 - \alpha_2 + \alpha_3) \beta + \alpha_2 \gamma) + (\alpha_4 \beta + \alpha_2 \gamma) \rho}{(2 \beta - \gamma)(\beta + \gamma)^2 + \gamma \rho^2}, \\
e_{22} &= \frac{- (2 \beta - \gamma)(\beta + \gamma) (\alpha_4 \beta + \alpha_2 \gamma) + \gamma ((\alpha_1 - \alpha_2 + \alpha_3) \beta + \alpha_2 \gamma) \rho}{(\beta - \gamma) (2 \beta - \gamma)(\beta + \gamma)^2 + \gamma \rho^2}, \\
e_{31} &= \frac{- (\beta - \gamma) (\beta + \gamma)^2 ((\alpha_1 + \alpha_2 - \alpha_3) \beta + \alpha_3 \gamma) + \beta (\beta + \gamma) (\alpha_4 \beta + \alpha_2 \gamma) \rho + \gamma (\alpha_1 \beta + \alpha_3 \gamma) \rho^2}{(\beta - \gamma)(\beta + \gamma) (2 \beta - \gamma)(\beta + \gamma)^2 + \gamma \rho^2}, \\
e_{42} &= \frac{- (2 \beta - \gamma)(\beta + \gamma) (\alpha_2 \beta + \alpha_4 \gamma) - \beta ((\alpha_1 - \alpha_2 + \alpha_3) \beta + \alpha_2 \gamma) \rho + \alpha_4 (\beta - \gamma) \rho^2}{(\beta - \gamma) (2 \beta - \gamma)(\beta + \gamma)^2 + \gamma \rho^2}.
\end{align*}
\]
Figure A.1: (Top left panel) The bipartite collaboration network $G$ of authors and projects analyzed in Example 2, where round circles represent authors and squares represent projects. (Top right panel) The projection of the bipartite network $G$ on the set of coauthors. The effort levels of the individual agents for each project they are involved in are indicated next to the nodes. (Bottom panel) The line graph $L(G)$ of Equation (6) associated with the collaboration network $G$, in which each node represents the effort an author invests into different projects. Solid lines indicate nodes sharing a project while dashed lines indicate nodes with the same author.
Next, we compute the above equilibrium effort levels using the equilibrium characterization in Equation (4). Note that \( d = (d_i)_{1 \leq i \leq 2} = (1, 2, 1, 1) \), \( n = (n_s)_{1 \leq s \leq 2} = (2, 2) \),
\[
\delta = (\delta_{is})_{1 \leq i \leq 3, 1 \leq s \leq 2} = \begin{pmatrix} 1 & 0 \\ 1 & 1 \\ 1 & 0 \\ 0 & 1 \end{pmatrix},
\]
and
\[
\varphi = (\varphi_{is})_{1 \leq i \leq 3, 1 \leq s \leq 2} = \begin{pmatrix} \frac{\beta \rho}{(\beta + \gamma)(2\beta - \gamma + \rho)} & 0 & 0 \\ \frac{\beta \rho}{(\beta - \gamma + \rho)(\beta + \gamma + \rho)} & \frac{\beta \rho}{(\beta + \gamma)(2\beta - \gamma + \rho)} & 0 \\ 0 & 0 & \frac{\beta \rho}{(\beta + \gamma)(\beta - \gamma + \rho)} \end{pmatrix}.
\]
Further, we have that \( \alpha = (\alpha_1, \alpha_2, \alpha_3, \alpha_4)^\top \) and
\[
\mu(\alpha) = \begin{pmatrix} -\frac{(\alpha_1 + \alpha_2 + \alpha_3)(\beta + \gamma)^2 + \alpha_2(\beta + \gamma)\rho + (\alpha_1 + \alpha_3)\gamma^2}{(\beta + \gamma)(2\beta - \gamma + \rho)(\beta + \gamma + \rho)} \\ \frac{\alpha_1 - 1 + \frac{\beta \rho}{\gamma^2}}{\alpha_2 - 1 + \frac{\beta \rho}{\gamma^2}} \end{pmatrix}.
\]
Next, we have that
\[
\Omega = \begin{pmatrix} -\frac{\beta \rho(3\beta + \gamma) + 2\rho}{(\beta + \gamma)(2\beta - \gamma + \rho)(\beta + \gamma + \rho)} & -\frac{\beta \rho}{(\beta + \gamma)(2\beta - \gamma + \rho)(\beta + \gamma + \rho)} \\ \frac{\beta \rho}{(\beta - \gamma + \rho)(\beta + \gamma + \rho)} & \frac{\beta \rho(2\beta - \gamma + \rho)}{(\beta + \gamma)(\beta - \gamma + \rho)(\beta + \gamma + \rho)} \end{pmatrix},
\]
and hence
\[
\epsilon = (I_2 - \Omega)^{-1} \mu(\alpha) = \begin{pmatrix} \frac{-(\alpha_1 + \alpha_2 + \alpha_3)(\beta - \gamma)(\beta + \gamma)^2 + (\alpha_4 \beta + \alpha_2 \gamma)\rho + (\alpha_1 + \alpha_3)\gamma^2}{(\beta - \gamma)(2\beta - \gamma + \rho)^2 + \gamma^2 \rho^2} \\ \frac{-(\alpha_2 + \alpha_4)(2\beta - \gamma - \beta + \gamma)^2 - (\beta + \gamma)(\alpha_1 - \alpha_2 + \alpha_3)(\beta + \gamma)\rho + (\alpha_4(\beta - \gamma) \rho^2}{(\beta - \gamma)(2\beta - \gamma + \rho)^2 + \gamma^2 \rho^2} \end{pmatrix}.
\]
Inserting the above expressions into Equation (4) yields exactly the equilibrium effort levels of Equation (30).

We now compute the equilibrium following Proposition 2. The matrix \( \Gamma \) with elements \( \Gamma_{is,jk} \) from Equation (6) can be written as follows
\[
\Gamma = \begin{pmatrix} 0 & \Gamma_{11,12} & \Gamma_{11,21} & \Gamma_{11,22} & \Gamma_{11,31} & \Gamma_{11,32} & \Gamma_{11,41} & \Gamma_{11,42} \\ \Gamma_{12,11} & 0 & \Gamma_{12,21} & \Gamma_{12,22} & \Gamma_{12,31} & \Gamma_{12,32} & \Gamma_{12,41} & \Gamma_{12,42} \\ \Gamma_{21,11} & \Gamma_{21,12} & 0 & \Gamma_{21,22} & \Gamma_{21,31} & \Gamma_{21,32} & \Gamma_{21,41} & \Gamma_{21,42} \\ \Gamma_{22,11} & \Gamma_{22,12} & \Gamma_{22,22} & 0 & \Gamma_{22,32} & \Gamma_{22,41} & \Gamma_{22,42} \\ \Gamma_{31,11} & \Gamma_{31,12} & \Gamma_{31,21} & \Gamma_{31,22} & 0 & \Gamma_{31,41} & \Gamma_{31,42} \\ \Gamma_{32,11} & \Gamma_{32,12} & \Gamma_{32,21} & \Gamma_{32,22} & \Gamma_{32,31} & 0 & \Gamma_{32,41} & \Gamma_{32,42} \\ \Gamma_{41,11} & \Gamma_{41,12} & \Gamma_{41,21} & \Gamma_{41,22} & \Gamma_{41,31} & \Gamma_{41,32} & 0 & \Gamma_{41,42} \\ \Gamma_{42,11} & \Gamma_{42,12} & \Gamma_{42,21} & \Gamma_{42,22} & \Gamma_{42,31} & \Gamma_{42,32} & \Gamma_{42,41} & 0 \end{pmatrix},
\]
\[
= \begin{pmatrix} 0 & 0 & -\beta & 0 & -\beta & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\beta & 0 & 0 & -\beta & 0 & 0 & 0 & 0 \\ 0 & 0 & -\beta & 0 & 0 & 0 & 0 & 0 \\ -\beta & -\beta & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.
\]
The FOC of Equation (63) in the proof of Proposition 2 in Appendix D can then be written as

\[
\begin{pmatrix}
\gamma & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \gamma & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \gamma & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \gamma & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \gamma
\end{pmatrix}
\begin{pmatrix}
\epsilon_{11} \\
\epsilon_{12} \\
\epsilon_{21} \\
\epsilon_{22} \\
\epsilon_{31} \\
\epsilon_{32} \\
\epsilon_{41} \\
\epsilon_{42}
\end{pmatrix}
= \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4
\end{pmatrix},
\]

from which we find that

\[
\begin{pmatrix}
\gamma \epsilon_{11} - \beta \epsilon_{21} - \beta \epsilon_{31} \\
\gamma \epsilon_{21} + \rho \epsilon_{22} - \beta \epsilon_{11} - \beta \epsilon_{31} \\
\gamma \epsilon_{22} + \rho \epsilon_{21} - \beta \epsilon_{21} \\
\gamma \epsilon_{31} - \beta \epsilon_{11} - \beta \epsilon_{21} \\
\gamma \epsilon_{42} - \beta \epsilon_{22}
\end{pmatrix}
= \begin{pmatrix}
\alpha_1 \\
\alpha_2 \\
\alpha_3 \\
\alpha_4
\end{pmatrix}.
\]

These are exactly the FOC from Equation (29).

An example of the equilibrium effort levels for agents 3 and 4 is shown in Figure A.2. The figure illustrates that with differently skilled agents, the effort of an agent in one project might exceed the effort in another project when the complementarity parameter \( \beta \) increases.

Example 3. Consider a network with 3 projects and 3 agents, where in the first project agents 1 and 2 are collaborating, in the second project agents 1 and 3 are collaborating and in the third project agents 2 and 3 are collaborating. An illustration can be found in Figure A.3. The payoffs
Figure A.3: (Top left panel) The bipartite collaboration network $G$ of authors and projects analyzed in Example 3, where round circles represent authors and squares represent projects. (Top right panel) The projection of the bipartite network $G$ on the set of coauthors. The effort levels of the individual agents for each project they are involved in are indicated next to the nodes. (Bottom panel) The line graph $L(G)$ of Equation (6) associated with the collaboration network $G$, in which each node represents the effort an author invests into different projects. Solid lines indicate nodes sharing a project while dashed lines indicate nodes with the same author.
of the agents are given by

\[
\begin{align*}
\pi_1 & = \epsilon_{11} \left( \alpha_1 + \frac{\beta}{2} \epsilon_{21} \right) + \epsilon_{21} \left( \alpha_2 + \frac{\beta}{2} \epsilon_{11} \right) + \epsilon_{12} \left( \alpha_1 + \frac{\beta}{2} \epsilon_{32} \right) + \epsilon_{32} \left( \alpha_3 + \frac{\beta}{2} \epsilon_{12} \right) \\
& \quad - \frac{\gamma}{2} \epsilon_{11}^2 - \frac{\gamma}{2} \epsilon_{12}^2 - \rho \epsilon_{11} \epsilon_{12} \\
\pi_2 & = \epsilon_{11} \left( \alpha_1 + \frac{\beta}{2} \epsilon_{21} \right) + \epsilon_{21} \left( \alpha_2 + \frac{\beta}{2} \epsilon_{11} \right) + \epsilon_{23} \left( \alpha_2 + \frac{\beta}{2} \epsilon_{33} \right) + \epsilon_{33} \left( \alpha_3 + \frac{\beta}{2} \epsilon_{23} \right) \\
& \quad - \frac{\gamma}{2} \epsilon_{21}^2 - \frac{\gamma}{2} \epsilon_{23}^2 - \rho \epsilon_{21} \epsilon_{23} \\
\pi_3 & = \epsilon_{32} \left( \alpha_3 + \frac{\beta}{2} \epsilon_{12} \right) + \epsilon_{12} \left( \alpha_1 + \frac{\beta}{2} \epsilon_{32} \right) + \epsilon_{33} \left( \alpha_3 + \frac{\beta}{2} \epsilon_{21} \right) + \epsilon_{21} \left( \alpha_2 + \frac{\beta}{2} \epsilon_{33} \right) \\
& \quad - \frac{\gamma}{2} \epsilon_{32}^2 - \frac{\gamma}{2} \epsilon_{33}^2 - \rho \epsilon_{32} \epsilon_{33}.
\end{align*}
\]

The first order conditions are given by

\[
\begin{align*}
\frac{\partial \pi_1}{\partial \epsilon_{11}} & = \alpha_1 + \epsilon_{21} \beta - \epsilon_{11} \gamma - \epsilon_{12} \rho = 0 \\
\frac{\partial \pi_1}{\partial \epsilon_{12}} & = \alpha_1 + \epsilon_{32} \beta - \epsilon_{12} \gamma - \epsilon_{11} \rho = 0 \\
\frac{\partial \pi_2}{\partial \epsilon_{21}} & = \alpha_2 + \epsilon_{11} \beta - \epsilon_{21} \gamma - \epsilon_{23} \rho = 0 \\
\frac{\partial \pi_2}{\partial \epsilon_{23}} & = \alpha_2 + \epsilon_{33} \beta - \epsilon_{23} \gamma - \epsilon_{21} \rho = 0 \\
\frac{\partial \pi_3}{\partial \epsilon_{32}} & = \alpha_3 + \epsilon_{12} \beta - \epsilon_{32} \gamma - \epsilon_{33} \rho = 0 \\
\frac{\partial \pi_3}{\partial \epsilon_{33}} & = \alpha_3 + \epsilon_{23} \beta - \epsilon_{33} \gamma - \epsilon_{32} \rho = 0. \tag{31}
\end{align*}
\]

Solving this system of equations directly yields

\[
\begin{align*}
\epsilon_{11} & = -\frac{(\beta - \gamma)(\alpha_2 \beta + \alpha_1 \gamma) + \alpha_3 \beta \rho + \alpha_1 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}, \\
\epsilon_{12} & = -\frac{(\beta - \gamma)(\alpha_3 \beta + \alpha_1 \gamma) + \alpha_2 \beta \rho + \alpha_1 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}, \\
\epsilon_{21} & = -\frac{(\beta - \gamma)(\alpha_1 \beta + \alpha_2 \gamma) + \alpha_3 \beta \rho + \alpha_2 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}, \\
\epsilon_{23} & = -\frac{(\beta - \gamma)(\alpha_3 \beta + \alpha_2 \gamma) + \alpha_1 \beta \rho + \alpha_2 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}, \\
\epsilon_{32} & = -\frac{(\beta - \gamma)(\alpha_1 \beta + \alpha_3 \gamma) + \alpha_2 \beta \rho + \alpha_3 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}, \\
\epsilon_{33} & = -\frac{(\beta - \gamma)(\alpha_2 \beta + \alpha_3 \gamma) + \alpha_1 \beta \rho + \alpha_3 \rho^2}{(\beta - \gamma - \rho)(\beta^2 - \gamma^2 + \beta \rho + \rho^2)}. \tag{32}
\end{align*}
\]

Next, we compute the above equilibrium effort levels using the equilibrium characterization in
Equation (4). Note that \( d = (d_i)_{1 \leq i \leq 3} = (2, 2, 2)^\top \), \( n = (n_s)_{1 \leq s \leq 3} = (2, 2, 2)^\top \),

\[
\delta = (\delta_{is})_{1 \leq i \leq 3, 1 \leq s \leq 3} = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{pmatrix},
\]

and

\[
\varphi = (\varphi_{is})_{1 \leq i \leq 3, 1 \leq s \leq 3} = \begin{pmatrix} \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & 0 \\ \frac{\beta \rho}{(\beta+\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta+\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta+\gamma+p)(\beta+\gamma+p)} \\ 0 & \frac{\beta \rho}{(\beta+\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta+\gamma+p)(\beta+\gamma+p)} \end{pmatrix}.
\]

Further, we have that \( \alpha = (\alpha_1, \alpha_2, \alpha_3)^\top \) and

\[
\mu(\alpha) = \begin{pmatrix} -\frac{(\alpha_1+\alpha_2)(\beta+\gamma-p)}{(\beta-\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_1+\alpha_2)(\beta+\gamma-p)}{(\beta+\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_1+\alpha_2)(\beta+\gamma-p)}{(\beta+\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_1+\alpha_2)(\beta+\gamma-p)}{(\beta+\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_1+\alpha_3)(\beta+\gamma-p)}{(\beta-\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_1+\alpha_3)(\beta+\gamma-p)}{(\beta+\gamma+p)(\beta+\gamma+p)} \\ -\frac{(\alpha_2+\alpha_3)(\beta+\gamma-p)}{(\beta-\gamma+p)(\beta+\gamma+p)} \end{pmatrix}.
\]

Next, we have that

\[
\Omega = \begin{pmatrix} \frac{2\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} \\ \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{2\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} \\ \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} & \frac{2\beta \rho}{(\beta-\gamma+p)(\beta+\gamma+p)} \end{pmatrix},
\]

and hence

\[
\epsilon = (I_2 - \Omega)^{-1} \mu(\alpha) = \begin{pmatrix} -\frac{(\alpha_1+\alpha_2)(\beta-\gamma)(\beta+\gamma)+2\alpha_3 \beta \rho+(\alpha_1+\alpha_2)\rho^2}{(\beta-\gamma-p)(\beta^2-\gamma^2+3\beta \rho+p^2)} \\ -\frac{(\alpha_1+\alpha_3)(\beta-\gamma)(\beta+\gamma)+2\alpha_2 \beta \rho+(\alpha_1+\alpha_3)\rho^2}{(\beta-\gamma-p)(\beta^2-\gamma^2+3\beta \rho+p^2)} \\ -\frac{(\alpha_2+\alpha_3)(\beta-\gamma)(\beta+\gamma)+2\alpha_1 \beta \rho+(\alpha_2+\alpha_3)\rho^2}{(\beta-\gamma-p)(\beta^2-\gamma^2+3\beta \rho+p^2)} \end{pmatrix}.
\]

Inserting the above expressions into Equation (4) yields exactly the equilibrium effort levels of Equation (32).

We next compute the equilibrium following Proposition 2. The matrix \( \Gamma \) with elements \( \Gamma_{is,jk} \)
The FOC of Equation (6) can be written as follows

\[ \Gamma = \begin{pmatrix}
0 & \Gamma_{11,12} & \Gamma_{11,13} & \Gamma_{11,21} & \Gamma_{11,22} & \Gamma_{11,23} & \Gamma_{11,31} & \Gamma_{11,32} & \Gamma_{11,33} \\
\Gamma_{12,11} & 0 & \Gamma_{12,13} & \Gamma_{12,21} & \Gamma_{12,22} & \Gamma_{12,23} & \Gamma_{12,31} & \Gamma_{12,32} & \Gamma_{12,33} \\
\Gamma_{13,11} & \Gamma_{13,12} & 0 & \Gamma_{13,21} & \Gamma_{13,22} & \Gamma_{13,23} & \Gamma_{13,31} & \Gamma_{13,32} & \Gamma_{13,33} \\
\Gamma_{21,11} & \Gamma_{21,12} & \Gamma_{21,13} & 0 & \Gamma_{21,22} & \Gamma_{21,23} & \Gamma_{21,31} & \Gamma_{21,32} & \Gamma_{21,33} \\
\Gamma_{22,11} & \Gamma_{22,12} & \Gamma_{22,13} & \Gamma_{22,21} & 0 & \Gamma_{22,22} & \Gamma_{22,31} & \Gamma_{22,32} & \Gamma_{22,33} \\
\Gamma_{23,11} & \Gamma_{23,12} & \Gamma_{23,13} & \Gamma_{23,21} & \Gamma_{23,22} & 0 & \Gamma_{23,31} & \Gamma_{23,32} & \Gamma_{23,33} \\
\Gamma_{31,11} & \Gamma_{31,12} & \Gamma_{31,13} & \Gamma_{31,21} & \Gamma_{31,22} & \Gamma_{31,23} & 0 & \Gamma_{31,32} & \Gamma_{31,33} \\
\Gamma_{32,11} & \Gamma_{32,12} & \Gamma_{32,13} & \Gamma_{32,21} & \Gamma_{32,22} & \Gamma_{32,23} & \Gamma_{32,31} & 0 & \Gamma_{32,33} \\
\Gamma_{33,11} & \Gamma_{33,12} & \Gamma_{33,13} & \Gamma_{33,21} & \Gamma_{33,22} & \Gamma_{33,23} & \Gamma_{33,31} & \Gamma_{33,32} & 0 \\
\end{pmatrix} \]

\[ = \begin{pmatrix}
0 & \rho & 0 & -\beta & 0 & 0 & 0 & 0 & 0 \\
\rho & 0 & 0 & 0 & 0 & 0 & 0 & -\beta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\beta & 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & -\beta \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\beta & 0 & 0 & 0 & 0 & 0 & \rho & 0 \\
0 & 0 & 0 & 0 & 0 & -\beta & 0 & \rho & 0 \\
\end{pmatrix}. \]

The FOC of Equation (63) in the proof of Proposition 2 in Appendix D can then be written as

\[ \left( \begin{pmatrix}
\gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \gamma & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & \gamma & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & \gamma & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \gamma & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \gamma & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & \gamma & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \gamma \\
\end{pmatrix} + \begin{pmatrix}
0 & \rho & 0 & -\beta & 0 & 0 & 0 & 0 & 0 \\
\rho & 0 & 0 & 0 & 0 & 0 & 0 & -\beta & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
-\beta & 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & \rho & 0 & 0 & 0 & -\beta \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho & 0 \\
0 & -\beta & 0 & 0 & 0 & 0 & 0 & \rho & 0 \\
0 & 0 & 0 & 0 & 0 & -\beta & 0 & \rho & 0 \\
\end{pmatrix} \right) \begin{pmatrix}
\epsilon_{11} \\
\epsilon_{12} \\
\epsilon_{13} \\
\epsilon_{21} \\
\epsilon_{22} \\
\epsilon_{23} \\
\epsilon_{31} \\
\epsilon_{32} \\
\epsilon_{33} \end{pmatrix} = \begin{pmatrix}
\alpha_1 \\
\alpha_1 \\
\alpha_1 \\
\alpha_2 \\
\alpha_2 \\
\alpha_2 \\
\alpha_3 \\
\alpha_3 \\
\alpha_3 \end{pmatrix}, \]

from which we find that

\[ \begin{pmatrix}
\gamma \epsilon_{11} + \rho \epsilon_{12} - \beta \epsilon_{21} \\
\gamma \epsilon_{12} + \rho \epsilon_{11} - \beta \epsilon_{32} \\
\gamma \epsilon_{21} - \beta \epsilon_{11} + \rho \epsilon_{23} \\
\gamma \epsilon_{23} + \rho \epsilon_{21} - \beta \epsilon_{33} \\
\gamma \epsilon_{32} - \beta \epsilon_{12} + \rho \epsilon_{33} \\
\gamma \epsilon_{33} - \beta \epsilon_{23} + \rho \epsilon_{32} \\
\end{pmatrix} = \begin{pmatrix}
\alpha_1 \\
\alpha_1 \\
\alpha_2 \\
\alpha_2 \\
\alpha_3 \\
\alpha_3 \end{pmatrix}. \]

However, these are exactly the FOC from Equation (31).

The total equilibrium effort levels for projects 1, 2 and 3 are shown in Figure A.4 for varying values of \( \beta \) and \( \rho \).

B. Alternative Payoff Functions

In the this appendix we study various alternative payoff specifications. These are summarized in Table 7. The assumption of the returns for an author from participating in a project to be split equally among the participants of the project is similar to the models studied in Jackson and Wolinsky [1996]; Kandel and Lazear [1992]. The assumption of a convex separable cost is similar to the model studied in Adams [2006]. The introduction of a quadratic cost with substitutes or complements, depending on the sign of the parameters \( \phi_{ss'} \), is similar to Cohen-Cole et al. [2012], and it includes the case of a convex total cost as a special case when \( \phi_{ss'} = \gamma \),
Figure A.4: Total equilibrium effort levels, $e_{s+} = \sum_{i=1}^{p} e_{is} \delta_{is}$, for projects 1, 2 and 3 with $\gamma = 1$, $\alpha_1 = 0.25$, $\alpha_2 = 0.5$, $\alpha_3 = 0.75$ and varying values of $\beta$ in the left panel, and varying values of $\rho$ in the right panel.

and the case of a convex separable cost when $\phi_{s,s'} = \gamma \delta_{s,s'}$. For the special case of only two activities, a theoretical model is studied in Belhaj and Deroïan [2014], and an empirical analysis is provided in Liu [2014]. The assumption of a time constraint is similar to Baumann [2014]; Salonen [2016].

C. Convex Separable Costs

Let $\delta_{is} \in \{0, 1\}$ indicate whether $i$ is participating in project $s$. The payoff of author $i$ is then given by

$$
\pi_i(G,e) = \sum_{s=1}^{p} \left( Y_s(G,e) - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is}
$$

$$
= \sum_{s=1}^{p} \left( \sum_{j \in \mathcal{N}_s} \alpha_j e_{js} + \frac{\beta}{2} \sum_{j \in \mathcal{N}_s, k \in \mathcal{N}_s \setminus \{j\}} e_{js} e_{ks} - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is}
$$

$$
= \sum_{s=1}^{p} \left( \sum_{j \in \mathcal{N}_s} e_{js} \left( \alpha_j + \frac{\beta}{2} \sum_{k \in \mathcal{N}_s \setminus \{j\}} e_{ks} \right) - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is},
$$

where $n_s = |\mathcal{N}_s|$ is the number of authors participating in project $s$, and we have that $n_s = \sum_{i=1}^{n} \delta_{is}$.

**Proposition 3.** Let the payoff function for each author $i = 1, \ldots, n$ be given by Equation (33). Then the unique interior Nash equilibrium effort levels are given by

$$
e_{is} = \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{(\beta + \gamma)(\gamma - \beta(n_s - 1))} \sum_{j \in \mathcal{N}_s} \alpha_j.
$$

for each author $i = 1, \ldots, n$ and each project $s = 1, \ldots, p$.

Inserting effort levels from Equation (43) into the production function from Equation (1)
return independent of number of authors
number of authors

\[ \pi = \sum_{s=1}^{n} \left( Y_s - \frac{1}{2} \epsilon_s^2 \delta_{is} \right) \]

\[ \pi_i = \sum_{s=1}^{n} \left( Y_s - \frac{1}{2} \epsilon_s^2 \delta_{is} \right)^2 \]

\[ \pi_i = \sum_{s=1}^{n} \left( \frac{1}{2} Y_s \delta_{is} - \frac{1}{2} \epsilon_s^2 \delta_{is} \right) \]

\[ \pi_i = \sum_{s=1}^{n} \left( \frac{1}{2} Y_s \delta_{is} - \frac{1}{2} \epsilon_s^2 \delta_{is} \right)^2 \]

<table>
<thead>
<tr>
<th>convex separable cost</th>
<th>convex total cost</th>
<th>quadratic cost with substitutes/complements</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Equation (1)]</td>
<td>[Equation (2)]</td>
<td>[Equation (3)]</td>
</tr>
</tbody>
</table>

The case of a quadratic cost includes the case of a convex total cost as a special case when \( \phi_{s,s'} = \frac{1}{2} \epsilon_s \epsilon_{s'} \), and the case of a convex separable cost when \( \phi_{s,s'} = \epsilon_s \epsilon_{s'} \).

Table 7. The alternative payoff specifications, \( \pi_i \) for \( i = 1, \ldots, n \). The output \( Y_s \) of project \( s \) is given in Equation (1). The case of convex separable cost is studied in Section C. The case of a quadratic case is studied in Section 3. The case of a quadratic cost includes the case of a convex total cost as a special case when \( \phi_{s,s'} = \frac{1}{2} \epsilon_s \epsilon_{s'} \), and the case of a convex separable cost when \( \phi_{s,s'} = \epsilon_s \epsilon_{s'} \).
yields

\[ Y_s(G) = \sum_{i \in N_s} e_{is} \left[ \alpha_i + \frac{\beta}{2} \sum_{j \in N_s \setminus \{i\}} e_{js} \right] \]
\[ = \frac{1}{2} \sum_{i \in N_s} e_{is} (\alpha_i + \gamma e_{is}) \]
\[ = \frac{1}{2} \sum_{i \in N_s} (\alpha_i e_{is} + \gamma e_{is}^2) \]
\[ = \frac{1}{2} \sum_{i \in N_s} \alpha_i \left( \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{\gamma - \beta(n_s - 1)} \sum_{j \in N_s \setminus \{i\}} \frac{\alpha_j}{\beta + \gamma} \right) + \gamma \left( \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{\gamma - \beta(n_s - 1)} \sum_{j \in N_s \setminus \{i\}} \frac{\alpha_j}{\beta + \gamma} \right)^2. \]

We next assume that author \( i \) receives a research funding, \( \tilde{z}_i \geq 0 \), proportional to the output she generates. In the case of a convex separable cost, the payoff of author \( i \) is given by

\[ \pi_i(G, e) = \sum_{s=1}^{p} \left( Y_s(G, e_s) - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is} + \sum_{s=1}^{p} Y_s(G, e_s) \tilde{z}_i \delta_{is} \]
\[ = \sum_{s=1}^{p} \left( (1 + \tilde{z}_i) Y_s(G, e_s) - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is} \]
\[ = \sum_{s=1}^{p} \left( z_i Y_s(G, e_s) - \frac{\gamma}{2} e_{is}^2 \right) \delta_{is}, \tag{36} \]

where we have denoted by \( z_i \equiv 1 + \tilde{z}_i \geq 1 \). The FOC is given by

\[ \frac{\partial \pi_i}{\partial e_{is}} = \left( z_i \frac{\partial Y_s}{\partial e_{is}} - \gamma e_{is} \right) \delta_{is} = z_i \left( \alpha_i + \beta \sum_{j \in N_s \setminus \{i\}} e_{js} \right) \delta_{is} - \gamma e_{is} \delta_{is} = 0. \]

We can write this as

\[ e_{is} = \frac{z_i \alpha_i}{\gamma + \beta z_i} + \frac{\beta z_i}{\gamma + \beta z_i} \sum_{j \in N_s} e_{js}. \tag{37} \]

Summation over all \( i \in N_s \) and rearranging terms yields

\[ \sum_{j \in N_s} e_{js} = \frac{\sum_{k \in N_s} \frac{z_k \alpha_k}{\gamma + \beta z_k}}{1 - \beta \sum_{k \in N_s} \frac{z_k}{\gamma + \beta z_k}}. \]

Inserting into Equation (37) gives equilibrium effort levels in the second stage, taken the research funds \( z_i \) as given

\[ e_{is} = \frac{z_i}{\gamma + \beta z_i} \left( \alpha_i + \beta \sum_{k \in N_s} \frac{z_k \alpha_k}{\gamma + \beta z_k} \right). \]
Denoting by \( \eta(z) \equiv \frac{z}{\gamma + \beta z} \), we can write equilibrium effort levels as follows

\[
e_{is} = \eta(z_i) \left( \alpha_i + \beta \frac{\sum_{k \in \mathcal{N}_i} \eta(z_k)}{1 - \beta \sum_{k \in \mathcal{N}_i} \eta(z_k)} \right). \tag{38}
\]

From the FOC we further have that

\[
e_{is} = \frac{z_i \alpha_i}{\gamma} \beta z_i \sum_{j \in \mathcal{N}_i \setminus \{i\}} e_{js},
\]

so that

\[
\beta \sum_{j \in \mathcal{N}_i \setminus \{i\}} e_{js} = \frac{\gamma e_{is}}{z_i} - \alpha_i
\]

and the output of project \( s \) can be written as

\[
Y_s(G, z) = \sum_{i \in \mathcal{N}_s} e_{is} \left[ \alpha_i + \frac{\beta}{2} \sum_{j \in \mathcal{N}_i \setminus \{i\}} e_{js} \right] = \frac{1}{2} \sum_{i \in \mathcal{N}_s} e_{is} \left( \alpha_i + \frac{\gamma}{z_i} e_{is} \right) = \frac{1}{2} \sum_{i \in \mathcal{N}_s} \eta(z_i) \left( \alpha_i + \beta \frac{\sum_{k \in \mathcal{N}_i} \eta(z_k)}{1 - \beta \sum_{k \in \mathcal{N}_i} \eta(z_k)} \right) \left( \alpha_i + \frac{\gamma \eta(z_i)}{z_i} \left( \alpha_i + \beta \frac{\sum_{k \in \mathcal{N}_i} \eta(z_k)}{1 - \beta \sum_{k \in \mathcal{N}_i} \eta(z_k)} \right) \right).
\]

With the equilibrium efforts \( e_{is} \) as a function of the funding \( z_i \) in Equation (38) for all \( i = 1, \ldots, n \) the planner then has to solve the following problem

\[
\max_{z \in \mathbb{R}^n_{\geq 1}} \left\{ \sum_{s=1}^{p} Y_s(G, z) - \sum_{i=1}^{n} \sum_{s=1}^{p} \delta_{is} Y_s(G, z) \right\} = \max_{z \in \mathbb{R}^n_{\geq 1}} \left\{ \sum_{s=1}^{p} Y_s(G, z) - \sum_{i=1}^{n} \sum_{s=1}^{p} (z_i - 1) Y_s(G, z) \delta_{is} \right\} = \max_{z \in \mathbb{R}^n_{\geq 1}} \left\{ \sum_{i=1}^{n} \sum_{s=1}^{p} \frac{1}{n} Y_s(G, z) - (z_i - 1) \delta_{is} Y_s(G, z) \right\} = \max_{z \in \mathbb{R}^n_{\geq 0}} \left\{ \sum_{i=1}^{n} \sum_{s=1}^{p} \frac{1}{n} (z_i - 1) \delta_{is} Y_s(G, z - u) \right\}.
\]

We denote by \( z^* \) the solution of the above optimization problem.
D. Proofs

Proof of Proposition 3. The first order condition (FOC) wrt $e_{is}$ is given by\(^{24}\)

$$\frac{\partial \pi_i(G, e)}{\partial e_{is}} = \sum_{s'=1}^{p} \left( \frac{\partial Y_{s'}(G, e_{s'})}{\partial e_{is}} - \gamma e_{is'} \right) \delta_{is'} = \left( \alpha_i + \beta \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js} \right) \delta_{is} - \gamma e_{is} \delta_{is} = 0, \quad (39)$$

where we have used the fact that

$$\frac{\partial Y_{s'}(G, e_{s'})}{\partial e_{is}} = \begin{cases} \alpha_i + \beta \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js}, & \text{if } s = s', \\ 0, & \text{otherwise}. \end{cases}$$

From Equation (52) we get

$$e_{is} = \frac{\alpha_i}{\gamma} + \frac{\beta}{\gamma} \sum_{j \in \mathcal{N}_s \setminus \{i\}} e_{js}, \quad (40)$$

for all projects $s$ in which $i$ is participating. Further, Equation (40) can be written as

$$\left( 1 + \frac{\beta}{\gamma} \right) e_{is} = \frac{\alpha_i}{\gamma} + \frac{\beta}{\gamma} \sum_{j \in \mathcal{N}_s} e_{js}, \quad (41)$$

and

$$e_{is} = \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{\beta + \gamma} \sum_{j \in \mathcal{N}_s} e_{js}. \quad (42)$$

Summation over $i \in \mathcal{N}_s$ gives

$$\sum_{j \in \mathcal{N}_s} e_{js} = \frac{1}{\beta + \gamma} \sum_{j \in \mathcal{N}_s} \alpha_j + \frac{\beta n_s}{\beta + \gamma} \sum_{j \in \mathcal{N}_s} e_{js},$$

and we get

$$\left( 1 - \frac{\beta n_s}{\beta + \gamma} \right) \sum_{j \in \mathcal{N}_s} e_{js} = \frac{1}{\beta + \gamma} \sum_{j \in \mathcal{N}_s} \alpha_j. \quad (42)$$

Hence

$$\sum_{j \in \mathcal{N}_s} e_{js} = \frac{1}{\gamma - \beta(n_s - 1)} \sum_{j \in \mathcal{N}_s} \alpha_j.$$ 

Inserting into Equation (42) yields

$$e_{is} = \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{(\beta + \gamma)(\gamma - \beta(n_s - 1))} \sum_{j \in \mathcal{N}_s} \alpha_j. \quad (43)$$

This allows us to determine the individual effort $e_{is}$ of author $i$ in project $s$. Denoting by $\tilde{\alpha}_i \equiv \frac{\alpha_i}{\beta + \gamma}$ we can write Equation (43) as

$$e_{is} = \tilde{\alpha}_i + \frac{\beta}{\gamma - \beta(n_s - 1)} \sum_{j \in \mathcal{N}_s} \tilde{\alpha}_j. \quad (44)$$

\(^{24}\)Observe that the second order condition (SOC) is given by $\frac{\partial^2 \pi_i(G, e)}{\partial e_{is}^2} = -\gamma \delta_{is} \leq 0.$
Further, denoting by \( \tilde{\beta}_s = \frac{\beta}{\gamma - \beta (n_s - 1)} \) this can be simplified to

\[
e_{is} = \tilde{\alpha}_i + \tilde{\beta}_s \sum_{j \in N_s} \tilde{\alpha}_j.
\]

(45)

From Equation (40) we have that

\[
\beta \sum_{j \in N_s \setminus \{i\}} e_{js} = \gamma e_{is} - \alpha_i,
\]

Inserting into the production function from Equation (1) yields

\[
Y_s(G) = \sum_{i \in N_s} e_{is} \left[ \alpha_i + \frac{\beta}{2} \sum_{j \in N_s \setminus \{i\}} e_{js} \right]
\]

\[
= \frac{1}{2} \sum_{i \in N_s} e_{is} \left( \alpha_i + \gamma e_{is} \right)
\]

\[
= \frac{1}{2} \sum_{i \in N_s} \left( \alpha_i e_{is} + \gamma e_{is}^2 \right)
\]

\[
= \frac{1}{2} \sum_{i \in N_s} \left( \alpha_i \left( \tilde{\alpha}_i + \tilde{\beta}_s \sum_{j \in N_s} \tilde{\alpha}_j \right) + \left( \tilde{\alpha}_i + \tilde{\beta}_s \sum_{j \in N_s} \tilde{\alpha}_j \right)^2 \right)
\]

\[
= \frac{1}{2} \sum_{i \in N_s} \left[ \alpha_i \left( \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{\gamma - \beta (n_s - 1)} \sum_{j \in N_s} \frac{\alpha_j}{\beta + \gamma} \right) + \gamma \left( \frac{\alpha_i}{\beta + \gamma} + \frac{\beta}{\gamma - \beta (n_s - 1)} \sum_{j \in N_s} \frac{\alpha_j}{\beta + \gamma} \right)^2 \right].
\]

(46)

\[\square\]

**Proof of Proposition 1.** The first order condition (FOC) wrt. \( e_{is} \) is given by\(^{25} \)

\[
\frac{\partial \pi_i(G,e)}{\partial e_{is}} = \sum_{s' = 1}^p \delta_{is'} \frac{\partial Y'_{s'}(G,e)}{\partial e_{is}} - \delta_{is} \sum_{s' = 1}^p \phi_{s,s'} e_{is'} \delta_{is'} = \left( \alpha_i + \beta \sum_{j \in N_s \setminus \{i\}} e_{js} \right) \delta_{is} - \delta_{is} \sum_{s' = 1}^p \phi_{s,s'} e_{is'} \delta_{is'} = 0,
\]

where we have used the fact that

\[
\frac{\partial Y'_{s'}(G,e_{s'})}{\partial e_{is}} = \begin{cases} 
\alpha_i + \beta \sum_{j \in N_s \setminus \{i\}} e_{js}, & \text{if } s = s', \\
0, & \text{otherwise}.
\end{cases}
\]

(47)

From Equation (47) we get

\[
\sum_{s' = 1}^p e_{is'} \phi_{s',s} \delta_{is} \delta_{is'} = \alpha_i \delta_{is} + \beta \sum_{j \in N_s \setminus \{i\}} e_{js} \delta_{js} \delta_{is},
\]

(48)

\[^{25}\text{Observe that the second order condition (SOC) is given by } \frac{\partial^2 \pi_i(G,e)}{\partial e_{is}^2} = -\phi_{s,s} \delta_{is} \leq 0.\]
for \( i = 1, \ldots, n \) and \( s = 1, \ldots, p \). In the following we denote by

\[
\hat{e}_{is} = \begin{cases} 
  e_{is}, & \text{if } i \in \mathcal{N}_s, \\
  0, & \text{otherwise.} 
\end{cases}
\] (49)

That is, we define \( \hat{e}_{is} \equiv \delta_{is} e_{is} \). Then we can write

\[
\delta_{is} \sum_{s' = 1}^{p} \hat{e}_{is'} \phi_{s's} = \alpha_i \delta_{is} + \beta \delta_{is} \sum_{j \in \mathcal{N}_s \setminus \{i\}} \hat{e}_{js},
\] (50)

In the following we assume that

\[
\phi_{ss'} = \begin{cases} 
  \gamma, & \text{if } s' = s, \\
  \rho, & \text{otherwise.}
\end{cases}
\]

Then we obtain from Equation (50) that

\[
\gamma \hat{e}_{is} + \rho \delta_{is} \sum_{s' = 1}^{p} \hat{e}_{is'} = \alpha_i \delta_{is} + \beta \delta_{is} \sum_{j \in \mathcal{N}_s \setminus \{i\}} \hat{e}_{js},
\] (51)

which can be written as follows

\[
(\gamma - \rho) \hat{e}_{is} + \rho \delta_{is} \sum_{s' = 1}^{p} \hat{e}_{is'} = \alpha_i \delta_{is} + \beta \delta_{is} \sum_{j \in \mathcal{N}_s \setminus \{i\}} \hat{e}_{js}.
\]

We can write this as

\[
(\beta + \gamma - \rho) \hat{e}_{is} + \rho \delta_{is} \sum_{s' = 1}^{p} \hat{e}_{is'} = \alpha_i \delta_{is} + \beta \delta_{is} \sum_{j \in \mathcal{N}_s \setminus \{i\}} \hat{e}_{js},
\]

and hence

\[
\delta_{is} \hat{e}_{+s} = \frac{\beta + \gamma - \rho}{\beta} \hat{e}_{is} - \frac{\alpha_i}{\beta} \delta_{is} + \frac{\rho}{\beta} \delta_{is} \hat{e}_{i+},
\] (52)

where we have denoted by

\[
\hat{e}_{+s} = \sum_{j \in \mathcal{N}_s} \hat{e}_{js},
\]

\[
\hat{e}_{i+} = \sum_{s = 1}^{p} \hat{e}_{is}.
\]

Summing over all \( i \in \mathcal{N}_s \) yields

\[
n_s \hat{e}_{+s} = \frac{\beta + \gamma - \rho}{\beta} \hat{e}_{+s} - \frac{1}{\beta} \sum_{i \in \mathcal{N}_s} \alpha_i \delta_{is} + \frac{\rho}{\beta} \sum_{i \in \mathcal{N}_s} \hat{e}_{i+}.
\]

Solving for \( \hat{e}_{+s} \) gives

\[
\hat{e}_{+s} = \frac{\rho}{\beta(n_s - 1) + \rho - \gamma} \sum_{i \in \mathcal{N}_s} \hat{e}_{i+} - \frac{1}{\beta(n_s - 1) + \rho - \gamma} \sum_{i \in \mathcal{N}_s} \alpha_i. \] (53)
Next, summation over all projects $s$ involving author $i$ in Equation (52) yields

$$\sum_{s=1}^{p} \delta_{is} \tilde{e}_{i+s} = \frac{\beta + \gamma - \rho}{\beta} \sum_{s=1}^{p} \tilde{e}_{is} - \frac{\alpha_i}{\beta} \sum_{s=1}^{p} \delta_{is} + \frac{\rho}{\beta} \tilde{e}_{i+} \sum_{s=1}^{p} \delta_{is},$$

and denoting by

$$d_i = \sum_{s=1}^{p} \delta_{is},$$

we get

$$\sum_{s=1}^{p} \delta_{is} \tilde{e}_{i+s} = \frac{\beta + \gamma + \rho(d_i - 1)}{\beta} \tilde{e}_{i+} - \frac{1}{\beta} \alpha_i d_i. \tag{54}$$

Solving for $\tilde{e}_{i+}$ gives

$$\tilde{e}_{i+} = \frac{\beta}{\beta + \gamma + \rho(d_i - 1)} \sum_{s=1}^{p} \delta_{is} \tilde{e}_{i+s} + \frac{1}{\beta + \gamma + \rho(d_i - 1)} \alpha_i d_i.$$

Summation over all $i \in N_s$ yields

$$\sum_{i \in N_s} \tilde{e}_{i+} = \sum_{i \in N_s} \beta \sum_{s'=1}^{p} \delta_{is'} \tilde{e}_{i+s'} + \sum_{i \in N_s} \frac{1}{\beta + \gamma + \rho(d_i - 1)} \alpha_i d_i. \tag{55}$$

Inserting Equation (55) into Equation (53) gives

$$\tilde{e}_{i+s} = \sum_{i=1}^{n} (\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1)) \sum_{s'=1}^{p} \delta_{is'} \tilde{e}_{i+s'}$$

$$+ \sum_{i=1}^{n} \frac{\rho \alpha_i d_i \delta_{is}}{(\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1))}$$

$$- \frac{1}{\beta(n_s - 1) + \rho - \gamma} \sum_{i=1}^{n} \alpha_i \delta_{is}. \tag{56}$$

In the following we denote by

$$\varphi_{is} \equiv \frac{\rho \beta \delta_{is}}{(\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1))},$$

$$\mu_s(\alpha) \equiv \sum_{i=1}^{n} \frac{\rho \alpha_i d_i \delta_{is}}{(\beta(n_s - 1) + \rho - \gamma)(\beta + \gamma + \rho(d_i - 1))} - \frac{1}{\beta(n_s - 1) + \rho - \gamma} \sum_{i=1}^{n} \alpha_i \delta_{is}. \tag{56}$$
Then we can write Equation (56) as follows

\[ \tilde{e}_{i+s} = \sum_{i=1}^{n} \varphi_{is} \sum_{s'=1}^{p} \delta_{is'} \tilde{e}_{s+s'} + \mu_s(\alpha) \]

\[ = \sum_{s'=1}^{p} \tilde{e}_{s+s'} \sum_{i=1}^{n} \varphi_{is} \delta_{is'} \]

\[ = \sum_{s'=1}^{p} \omega_{ss'} \tilde{e}_{s+s'} + \mu_s(\alpha), \]

where we have denoted by

\[ \omega_{ss'} \equiv \sum_{i=1}^{n} \varphi_{is} \delta_{is'}. \]

Further, let \( \epsilon \equiv (\tilde{e}_{i+1}, \ldots, \tilde{e}_{i+p})^\top \) and \( \Omega \equiv (\omega_{ss'})_{1 \leq s, s' \leq p} \), then we can write the above equation in vector-matrix form as

\[ \epsilon = \Omega \epsilon + \mu(\alpha). \]

That is

\[ (I_p - \Omega) \epsilon = \mu(\alpha). \]

When the matrix \( I_p - \Omega \) is invertible, then we can write

\[ \epsilon = (I_p - \Omega)^{-1} \mu(\alpha). \tag{57} \]

Next, inserting Equation (57) into Equation (54) gives

\[ \sum_{s=1}^{p} \delta_{is} \epsilon_s = \frac{\beta + \gamma + \rho(d_i - 1)}{\beta} \tilde{e}_{i+} - \frac{1}{\beta} \alpha_i d_i, \]

so that

\[ \tilde{e}_{i+} = \frac{\beta}{\beta + \gamma + \rho(d_i - 1)} \sum_{s=1}^{p} \delta_{is} \epsilon_s + \frac{1}{\beta + \gamma + \rho(d_i - 1)} \alpha_i d_i. \tag{58} \]

Moreover, note that Equation (52) can be written as

\[ \tilde{e}_{is} = \frac{\beta}{\beta + \gamma - \rho} \delta_{is} \tilde{e}_{i+s} + \frac{1}{\beta + \gamma - \rho} \alpha_i \delta_{is} - \frac{\rho}{\beta + \gamma - \rho} \delta_{is} \tilde{e}_{i+}. \tag{59} \]

Inserting Equations (58) and (57) into Equation (59) gives

\[ \tilde{e}_{is} = \frac{\beta}{\beta + \gamma - \rho} \delta_{is} \epsilon_s + \frac{1}{\beta + \gamma - \rho} \alpha_i \delta_{is} - \frac{\rho \beta}{(\beta + \gamma - \rho)(\beta + \gamma + \rho(d_i - 1))} \delta_{is} \sum_{s'=1}^{p} \epsilon_{s'} \delta_{is'} - \frac{\rho}{(\beta + \gamma - \rho)(\beta + \gamma + \rho(d_i - 1))} \delta_{is} \alpha_i d_i. \tag{60} \]
Equation (60) can be written as follows

\[
\hat{e}_{is} = \frac{\delta_{is}}{\beta + \gamma - \rho} \left[ \beta e_a + \alpha_i - \frac{\rho}{\beta + \gamma + \rho(d_i - 1)} \left( \sum_{s'=1}^{p} \delta_{is'} e_{s'} + \alpha_i d_i \right) \right].
\] (61)

\[\text{Proof of Proposition 2.}\] First note that the FOC of the effort levels can be written as (see Equation (51) in the proof of Proposition 1 in Appendix D)

\[
e_{is} \delta_{is} + \frac{\rho \delta_{is}}{\beta + \gamma - \rho} \sum_{s=1}^{p} e_{is} \delta_{is} - \frac{\beta \delta_{is}}{\beta + \gamma - \rho} \sum_{j \in N_s} e_{js} \delta_{js} = \frac{1}{\beta + \gamma - \rho} \alpha_i \delta_{is}.\] (62)

Then, introducing the \((n \times p) \times (n \times p)\) matrix

\[
\Gamma_{is,jk} = \begin{cases} 
\rho \delta_{is} \delta_{jk}, & \text{if } i = j, s \neq k, \\
-\beta \delta_{is} \delta_{jk}, & \text{if } i \neq j, s = k, \\
0, & \text{otherwise},
\end{cases}
\]

and let \(\delta\) be an \((n \times p)\) stacked vector with elements \(\delta_{is}\), \(\alpha\) an \((n \times p)\) stacked vector with elements \(\alpha_{is} = \alpha_i \delta_{is}\) and \(e\) an \((n \times p)\) stacked vector of agent-project effort levels, \(e_{is}\), we can write Equation (62) in vector-matrix notation compactly as follows

\[
(\gamma \ \text{diag}(\delta) + \Gamma) \ e = \alpha,
\] (63)

so that, when the matrix \(\gamma \ \text{diag}(\delta) + \Gamma\) is invertible, we can write the equilibrium effort levels as follows

\[
e = (\gamma \ \text{diag}(\delta) + \Gamma)^{-1} \ \alpha.
\]

\[\square\]

\[\text{E. Data Appendix}\]

We use the following variables, retrieved in January 2017:

- 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 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.\(^{26}\)
  4. Current RePEc ranking for the main affiliation for 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.

\(^{26}\)See [https://ideas.repec.org/top/top.person.all.html](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.
13. Gender as determined by a likelihood table using the first and possibly middle name. Uncertain matches were almost all resolved through internet search.
15. Fields of work, as determined by the NEP fields for which their working papers were selected for email dissemination.

• Potential author pair characteristics
  1. Student-advisor relationship, as recorded in the RePEc Genealogy.
  2. Joint alma mater of terminal decree as recorded in the RePEc Genealogy.
  3. 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.
  4. Joint ethnicity.
  5. Joint country of main affiliation.
  6. Joint field of work. There are two ways we determine this, both based on the NEP fields in which the authors published. For the first, we only consider the fields in which each author has written at least five papers or 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.

F. Sufficient Conditions for Existence and Uniqueness

Sufficient Condition 1: Let

\[ M_1 = \phi[I_{np} - D(\tau_p \tau_p' \otimes I_n)] - \beta[I_{np} - D(I_p \otimes \tau_n \tau_n')] \]

The matrix \( S = I_{np} - M_1 \) is nonsingular, if \( \|M_1\|_1 < 1 \), where \( \|\cdot\|_1 \) is the column sum matrix norm. As

\[
\|M_1\|_1 = \|\phi[I_{np} - D(\tau_p \tau_p' \otimes I_n)] - \beta[I_{np} - D(I_p \otimes \tau_n \tau_n')]\|_1 \\
\leq \|\phi\| \cdot \|I_{np} - D(\tau_p \tau_p' \otimes I_n)\|_1 + |\beta| \cdot \|I_{np} - D(I_p \otimes \tau_n \tau_n')\|_1 \\
\leq |\phi| (\|I_{np}\|_1 + \|D(\tau_p \tau_p' \otimes I_n)\|_1) + |\beta| (\|I_{np}\|_1 + \|D(I_p \otimes \tau_n \tau_n')\|_1) \\
\leq |\phi| (1 + \max_{i=1,\ldots,n} \sum_{s=1}^p \delta_{i,s}) + |\beta| (1 + \max_{i=1,\ldots,n} \sum_{s=1}^n \delta_{i,s}),
\]

a sufficient condition for the nonsingularity of \( S \) is

\[
|\phi| (1 + \max_{i=1,\ldots,n} \sum_{s=1}^p \delta_{i,s}) + |\beta| (1 + \max_{i=1,\ldots,n} \sum_{s=1}^n \delta_{i,s}) < 1. \tag{64}
\]

If (64) is satisfied, then \( S^{-1} \) can be expanded in polynomials of \( M_1 \):

\[
S^{-1} = I_{np} + M_1 + M_1^2 + \cdots.
\]

Then, we can use the leading order terms of the expansion (say, \( I_{np} + M_1 + M_1^2 \) or \( I_{np} + M_1 \)) to approximate \( S^{-1} \). This can be computationally more efficient for larger datasets. In the following we derive an alternative sufficient condition for the nonsingularity of the matrix \( S \).

Sufficient Condition 2: Suppose \( 1 + \beta - \phi \neq 0 \). Let

\[ M_2 = \beta^* D(I_p \otimes \tau_n \tau_n') - \phi^* D(\tau_p \tau_p' \otimes I_n), \]

where \( \beta^* = \beta/(1 + \beta - \phi) \) and \( \phi^* = \phi/(1 + \beta - \phi) \). The matrix \( S = I_{np} - M_2 \) is nonsingular, if \( \|M_2\|_1 < 1 \), where \( \|\cdot\|_1 \) is the column sum matrix norm. As

\[
\|M_2\|_1 = \|\beta^* D(I_p \otimes \tau_n \tau_n') - \phi^* D(\tau_p \tau_p' \otimes I_n)\|_1 \\
\leq |\beta^*| \cdot \|D(I_p \otimes \tau_n \tau_n')\|_1 + |\phi^*| \cdot \|D(\tau_p \tau_p' \otimes I_n)\|_1 \\
\leq |\beta^*| \max_{s=1,\ldots,p} \sum_{i=1}^n \delta_{i,s} + |\phi^*| \max_{i=1,\ldots,n} \sum_{s=1}^p \delta_{i,s},
\]

a sufficient condition for the nonsingularity of \( S \) is

\[
|\beta^*| \max_{s=1,\ldots,p} \sum_{i=1}^n \delta_{i,s} + |\phi^*| \max_{i=1,\ldots,n} \sum_{s=1}^p \delta_{i,s} < 1. \tag{65}
\]
If (65) is satisfied, then $S^{-1}$ can be expanded in polynomials of $M_2$:

$$S^{-1} = I_{np} + M_2 + M_2^2 + \cdots.$$ 

Then, we can use the leading order terms of the expansion (say, $I_{np} + M_2 + M_2^2$ or $I_{np} + M_2$) to approximate $S^{-1}$.

G. Heuristic Explanation of the Estimation Bias

In this appendix we provide a heuristic explanation on why the coefficients $\beta$ and $\phi$ estimated from Model (I) of Table 2 would be downward biased. First note that it is not straightforward to exploit data variations to gauge the direction of estimation bias because the variations of paper qualities, paper authorships, and author characteristics have been distorted nonlinearly in the equilibrium effort of Eq. (24) and the predicted output of Eq. (25). Alternatively, we consider a comparison between original and counterfactual scenarios. In the original scenario, we use the estimates of $\beta$, $\phi$, and $\lambda$ from Model (I) and Model (II), respectively, to predict papers’ outputs. Intuitively, these predicted values are the best approximation of the real values that each model can offer. In the counterfactorial scenario, we take the estimated authors’ abilities from Model (I), but manipulate the predicted research efforts and outputs by using the “higher” values of $\beta$ and $\phi$ from Model (II). The goal is to show that these manipulated outputs will be further apart from the real ones.

In Figure G.1 we contrast the author abilities obtained from Model (I) and from Model (II). Compared to Model (II), Model (I) overestimates the average value, but underestimates the dispersion. We believe that this is due to the fact that Model (I) omits the individual latent variables. Followed by this pattern, the computed equilibrium efforts from Model (I) also show a higher average but a smaller dispersion compared to Model (II), as shown in Figure G.2 (a). In addition, if we manipulate the computation of efforts from Model (I) by using the higher values of $\beta$ and $\phi$ from Model (II), Figure G.2 (b) shows that the efforts from Model (I) become even more concentrated and deviate further away from the result of Model (II). In Figure I.1 we continue to contrast the predicted paper outputs from Model (I) and Model (II). In panel (a), we can see that Model (II) provides better predictions than Model (I), despite that both Models fail to match a large amount of real values which cluster near zero. When we again manipulate the predicted paper outputs from Model (I) by using higher values of $\beta$ and $\phi$ from Model (II), the predictions of Model (I) become more concentrated at one middle range, which make the whole distribution further deviate from the true one. From these comparisons, we can conclude that due to omitting individual latent variables, Model (I) needs to underestimate the values of $\beta$ and $\phi$ in order to maintain a better goodness of fit to the real data.

H. Simulation Study

To show the proposed Bayesian MCMC estimation approach in Section 5.4 can effectively recover the true parameters from the model of Eqs. (26) and (27), we conduct a Monte Carlo simulation study to examine the bias and standard deviation from estimation results. The simulation consists of 100 repetitions. In each repetition, we first simulate dyadic binary exogenous variables $c_{i,s}$ by drawing two uniform random variables, $z_i$ and $z_s$. If both $z_i$ and $z_s$ are above 0.7 or below 0.3, we set $c_{i,s} = 1$; otherwise, we set $c_{i,s} = 0$. We simulate individual exogenous variables $x$, author latent variables $\mu$, and project latent variables $\kappa$ from standard normal distributions. Then we generate the artificial project output $y$ and participation $D$ based on the data generating process (DGP) in Eqs. (26) and (27). We estimate two models, one is the full
model (i.e., the DGP model) where both project output and project participation are endogenous and the other is just the project output equation by treating the collaboration matrix $D$ as exogenous. We conduct simulations with two sample sizes to show how data information can improve estimation accuracy in finite samples.

The simulation results are shown in Table 8. We report the bias and the standard deviation based on the point estimate of each coefficient across repetitions. First of all, we observe that when treating the collaboration network as exogenous, there are downward biases on the estimates of $\beta$ and $\phi$. This is the same problem that we can reproduce from our empirical findings, which strengthens our argument that when omitting individual latent variables, the variation on authors’ abilities will be underestimated and it results in lower estimates of $\beta$ and $\phi$. The second thing to be observed from the table is when using the full model, we mostly recover the true value of each coefficient, despite of small finite sample biases. However, these finite sample biases fade off when the sample size increases, which tells us that the proposed estimation algorithm has the desired finite sample performance.

I. Goodness-of-Fit Statistics

J. Heterogeneous Spillover on Authors’ Efforts

Suppose there are $n$ agents and $p$ projects. The production function of project $r$, $r = 1, \cdots, p$, is given by

$$y_r = \sum_{i=1}^{n} \delta_{i,r} e_{i,r} \xi_i + \beta \sum_{i=1}^{n} \sum_{j=1}^{n} f_{ij} \delta_{i,r} \delta_{j,r} e_{i,r} e_{j,r} + \epsilon_r,$$  

(66)
Figure G.2: Distributions of efforts computed from the models of exogenous and endogenous project matching
Figure G.3: Distributions of paper qualities computed from the models of exogenous and endogenous project matching
Figure I.1: Goodness-of-fit test for coauthorship network
where $\delta_{i,r}$ is an indicator variable such that $\delta_{i,r} = 1$ if agent $i$ participates project $r$ and $\delta_{i,r} = 0$ otherwise. $e_{i,r}$ is the effort of agent $i$ put into project $r$. $\xi_i = x_i \lambda$ is a measure of the productivity of agent $i$, and $\epsilon_r \sim (0, \sigma_r^2)$ is a project-specific random shock. $f_{ij}$ is an exogenous proximity measure between researchers $i$ and $j$. We set $f_{ii} = 0$. The econometrician observes the $y_r, x_i$ and $\delta_{i,r}$, but not $e_{i,r}$ and $\epsilon_r$.

The utility function of agent $i$ is given by

$$\pi_i = \sum_{r=1}^p \delta_{i,r} y_r \left( -\frac{1}{2} \sum_{r=1}^p \delta_{i,r} e_{i,r}^2 + \phi \sum_{r=1}^p \sum_{s=1,s \neq r}^{p} \delta_{i,r} \delta_{i,s} e_{i,r} e_{i,s} \right) + \lambda \sum_{j=1}^n f_{ij} \delta_{i,r} \delta_{j,r} e_{j,r} e_{k,r} + \epsilon_r \right) \right) - \frac{1}{2} \sum_{r=1}^p \delta_{i,r} \delta_{i,s} e_{i,s}. \right)$$

Substitution of (66) into (67) gives

$$\pi_i = \sum_{r=1}^p \delta_{i,r} \left( \sum_{j=1}^n f_{ij} \delta_{j,r} e_{j,r} \xi_j + \lambda \sum_{j=1}^n f_{ij} \delta_{j,r} \delta_{j,r} e_{j,r} e_{k,r} + \epsilon_r \right) - \frac{1}{2} \sum_{r=1}^p \delta_{i,r} \delta_{i,s} e_{i,s}. \right)$$

The FOC with respect to $e_{i,r}$ is given by

$$\frac{\partial \pi_i}{\partial e_{i,r}} = \delta_{i,r} \xi_i + \lambda \sum_{j=1}^n f_{ij} \delta_{j,r} e_{j,r} e_{j,r} + \epsilon_r \right) - \frac{1}{2} \sum_{r=1}^p \delta_{i,r} \delta_{i,s} e_{i,s} = 0,$$

which implies

$$\delta_{i,r} e_{i,r} = \delta_{i,r} \left( \xi_i + \lambda \sum_{j=1}^n f_{ij} \delta_{j,r} e_{j,r} + \phi \sum_{s=1,s \neq r}^{p} \delta_{i,r} \delta_{i,s} e_{i,s} \right), \tag{68}$$

or

$$\tilde{e}_{i,r} = \delta_{i,r} \left( \xi_i + \lambda \sum_{j=1}^n f_{ij} \tilde{e}_{j,r} + \phi \sum_{s=1,s \neq r}^{p} \tilde{e}_{i,s} \right), \tag{69}$$

where $\tilde{e}_{i,r} = \delta_{i,r} e_{i,r}$. Note that (22) can be rewritten as

$$\tilde{e}_{i,r} = \delta_{i,r} \left( \xi_i + \lambda \sum_{j=1}^n f_{ij} \tilde{e}_{j,r} + \phi \sum_{s=1,s \neq r}^{p} \tilde{e}_{i,s} \right). \tag{70}$$

Let $\xi = (\xi_1, \ldots, \xi_n)' = X \beta$ and $F = [f_{ij}]$. Let $\tilde{e}_r = (\tilde{e}_{1,r}, \ldots, \tilde{e}_{n,r})'$ and $\tilde{e} = (\tilde{e}_1', \ldots, \tilde{e}_p')'$. Then,
in vector form, \((22)\) becomes

\[
\tilde{e} = D[\iota_p \otimes \xi + \lambda(I_p \otimes F)\tilde{e} - \phi(A \otimes I_n)\tilde{e}]
\]

where \(D = \text{diag}^p_{r=1}\{\text{diag}^n_{i=1}\{\delta_{i,r}\}\}\), \(\iota_n\) is an \(n \times 1\) vector of ones, and \(A = \iota_p t_p' - I_p\). Hence, the equilibrium effort is given by

\[
\tilde{e}^* = [I_{np} - \lambda D(I_p \otimes F) + \phi D(A \otimes I_n)]^{-1}D(\iota_p \otimes \xi).
\]

(71)

**K. Estimation Results from other Sample Periods**

Table 9: Summary statistics for the 2007-2009 sample.

<table>
<thead>
<tr>
<th>Papers</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>S.D.</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation recursive Impact Factor</td>
<td>0.0000</td>
<td>91.2845</td>
<td>10.9982</td>
<td>17.8867</td>
<td>3601</td>
</tr>
<tr>
<td>number of authors (in each paper)</td>
<td>1</td>
<td>4</td>
<td>1.8298</td>
<td>0.7057</td>
<td>3601</td>
</tr>
</tbody>
</table>

**Authors**

| Log life-time citations         | 0       | 10.5394  | 5.7806  | 1.6097  | 1724        |
| Decades after Ph.D. graduation  | -0.8    | 5.30000  | 0.9959  | 0.9410  | 1724        |
| Female                           | 0       | 1        | 0.1340  | 0.3407  | 1724        |
| NBER connection                  | 0       | 1        | 0.1259  | 0.3318  | 1724        |
| Ivy League connection            | 0       | 1        | 0.1618  | 0.3684  | 1724        |
| Editor                           | 0       | 1        | 0.0574  | 0.2327  | 1724        |
| number of papers (for each author) | 1      | 49       | 3.8219  | 3.8321  | 1724        |

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.

Table 10: Summary statistics for the 2013-2015 sample.

<table>
<thead>
<tr>
<th>Papers</th>
<th>Min</th>
<th>Max</th>
<th>Mean</th>
<th>S.D.</th>
<th>Sample size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citation recursive Impact Factor</td>
<td>0.0000</td>
<td>43.1194</td>
<td>3.1582</td>
<td>1.8687</td>
<td>1941</td>
</tr>
<tr>
<td>number of authors (in each paper)</td>
<td>1</td>
<td>5</td>
<td>1.9619</td>
<td>0.6891</td>
<td>1941</td>
</tr>
</tbody>
</table>

**Authors**

| Log life-time citations         | 0       | 10.5394  | 5.2776  | 1.8687  | 1301        |
| Decades after Ph.D. graduation  | -0.2    | 10.2000  | 1.2784  | 1.0005  | 1301        |
| Female                           | 0       | 1        | 0.1253  | 0.3312  | 1301        |
| NBER connection                  | 0       | 1        | 0.1238  | 0.3294  | 1301        |
| Ivy League connection            | 0       | 1        | 0.1460  | 0.3533  | 1301        |
| Editor                           | 0       | 1        | 0.0507  | 0.2195  | 1301        |
| number of papers (for each author) | 1      | 52       | 2.9270  | 3.1572  | 1301        |

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.
Table 11: Estimation results for other Sample Periods

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Model (I)</td>
<td>Model (II)</td>
</tr>
<tr>
<td></td>
<td>Model (I)</td>
<td>Model (II)</td>
</tr>
<tr>
<td><strong>Output</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\beta$</td>
<td>-0.0692***</td>
<td>-0.0466</td>
</tr>
<tr>
<td></td>
<td>(0.0228)</td>
<td>(0.0443)</td>
</tr>
<tr>
<td>$\phi$</td>
<td>-0.0073</td>
<td>0.0015</td>
</tr>
<tr>
<td></td>
<td>(0.0042)</td>
<td>(0.0078)</td>
</tr>
<tr>
<td>Constant</td>
<td>-2.2244***</td>
<td>-0.9490***</td>
</tr>
<tr>
<td></td>
<td>(0.4534)</td>
<td>(0.2856)</td>
</tr>
<tr>
<td>Log life-time citations</td>
<td>0.7910***</td>
<td>0.4731***</td>
</tr>
<tr>
<td></td>
<td>(0.0638)</td>
<td>(0.0542)</td>
</tr>
<tr>
<td>Decades after graduation</td>
<td>1.8056***</td>
<td>-0.1793</td>
</tr>
<tr>
<td></td>
<td>(1.0958)</td>
<td>(0.4336)</td>
</tr>
<tr>
<td>(Decades after graduation)$^2$</td>
<td>-3.9824***</td>
<td>-1.3618</td>
</tr>
<tr>
<td></td>
<td>(1.3521)</td>
<td>(0.6057)</td>
</tr>
<tr>
<td>(Decades after graduation)$^3$</td>
<td>2.1675***</td>
<td>1.1673</td>
</tr>
<tr>
<td></td>
<td>(0.7336)</td>
<td>(0.3288)</td>
</tr>
<tr>
<td>(Decades after graduation)$^4$</td>
<td>-0.4764***</td>
<td>-0.3052**</td>
</tr>
<tr>
<td></td>
<td>(0.1734)</td>
<td>(0.0673)</td>
</tr>
<tr>
<td>(Decades after graduation)$^5$</td>
<td>0.0362**</td>
<td>0.0244*</td>
</tr>
<tr>
<td></td>
<td>(0.0146)</td>
<td>(0.0064)</td>
</tr>
<tr>
<td>Female</td>
<td>0.3994**</td>
<td>-0.2788</td>
</tr>
<tr>
<td></td>
<td>(0.2140)</td>
<td>(0.2654)</td>
</tr>
<tr>
<td>NBER connection</td>
<td>0.4219***</td>
<td>1.2190***</td>
</tr>
<tr>
<td></td>
<td>(0.1430)</td>
<td>(0.1254)</td>
</tr>
<tr>
<td>Ivy League connection</td>
<td>0.8790***</td>
<td>0.1295</td>
</tr>
<tr>
<td></td>
<td>(0.1322)</td>
<td>(0.1156)</td>
</tr>
<tr>
<td>Editor</td>
<td>-0.7919***</td>
<td>-0.3160</td>
</tr>
<tr>
<td></td>
<td>(0.2687)</td>
<td>(0.2160)</td>
</tr>
<tr>
<td>$\rho$</td>
<td>6.2949***</td>
<td>3.4989**</td>
</tr>
<tr>
<td></td>
<td>(3.3209)</td>
<td>(1.1559)</td>
</tr>
<tr>
<td>$\eta$</td>
<td>-0.0352</td>
<td>-0.6512</td>
</tr>
<tr>
<td></td>
<td>(0.9206)</td>
<td>(0.4871)</td>
</tr>
<tr>
<td>$\sigma^2$</td>
<td>235.4862***</td>
<td>152.8730***</td>
</tr>
<tr>
<td></td>
<td>(5.6207)</td>
<td>(3.8297)</td>
</tr>
<tr>
<td><strong>Matching</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>constant</td>
<td>-9.8859***</td>
<td>-9.0102***</td>
</tr>
<tr>
<td></td>
<td>(1.1301)</td>
<td>(1.1351)</td>
</tr>
<tr>
<td>Same NEP</td>
<td>0.4036***</td>
<td>1.2067**</td>
</tr>
<tr>
<td></td>
<td>(0.0510)</td>
<td>(0.0664)</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>0.1559</td>
<td>0.6788***</td>
</tr>
<tr>
<td></td>
<td>(0.0853)</td>
<td>(0.1083)</td>
</tr>
<tr>
<td>Affiliation</td>
<td>1.8167***</td>
<td>4.6835***</td>
</tr>
<tr>
<td></td>
<td>(0.1713)</td>
<td>(0.2298)</td>
</tr>
<tr>
<td>Female</td>
<td>-0.3007***</td>
<td>-0.0020</td>
</tr>
<tr>
<td></td>
<td>(0.1018)</td>
<td>(0.1236)</td>
</tr>
<tr>
<td>Advisor-advisee</td>
<td>1.3886***</td>
<td>5.1265***</td>
</tr>
<tr>
<td></td>
<td>(0.1119)</td>
<td>(0.1863)</td>
</tr>
<tr>
<td>Past coauthors</td>
<td>7.8126***</td>
<td>5.1480***</td>
</tr>
<tr>
<td></td>
<td>(1.051)</td>
<td>(1.1236)</td>
</tr>
<tr>
<td>Share common co-authors</td>
<td>2.0399***</td>
<td>2.2454***</td>
</tr>
<tr>
<td></td>
<td>(0.3542)</td>
<td>(0.2426)</td>
</tr>
<tr>
<td>Author effect</td>
<td>3.0220***</td>
<td>3.7976***</td>
</tr>
<tr>
<td></td>
<td>(0.1949)</td>
<td>(0.2097)</td>
</tr>
<tr>
<td>Project effect</td>
<td>0.0533</td>
<td>0.2687</td>
</tr>
<tr>
<td></td>
<td>(0.2951)</td>
<td>(0.2659)</td>
</tr>
<tr>
<td><strong>Sample size</strong></td>
<td>3601</td>
<td>1941</td>
</tr>
</tbody>
</table>

Note: Model (I): assume exogenous matching between authors and papers. Model (2): assume endogenous matching by Equation (26).