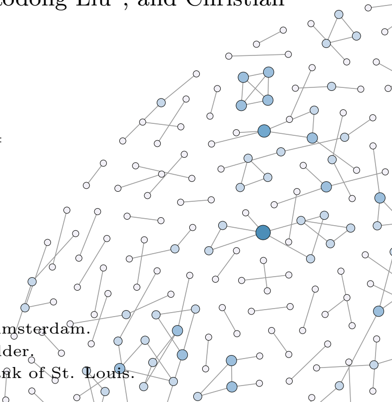
Collaboration in Bipartite Networks

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

EEA-ESEM-2024



[1]Department of Economics, National Taiwan University
[2]School of Business and Economics, Vrije Universiteit Amsterdam.
[3]Department of Economics, University of Colorado Boulder.
[4]Department of Economic Research, Federal Reserve Bank of St. Louis.

Motivation

- Collaboration between agents plays a prominent role in team production.
- For example, co-inventor networks in technology innovation (Singh, 2005; Fleming et al., 2007), co-authorship networks in scientific research (Ductor et al., 2014), corporate board members in firm performance (Conyon & Muldoon, 2006).
- The aim of this paper is to develop a general structural model that helps us to understand how collaboration affects team production.
- In our model, an agent's contribution to the team's project is determined by his/her effort. Individual agent choose **endogenous** effort on projects in a network game.
- Collaborations in our model bring in two possible effects on efforts: the **complementarity** effect and the **substitutability** effect due to congestion (Jackson and Wolinsky, 1996).

Overview

- Theoretical model
  - Network game of multiple agents who participate in overlapping projects.
  - We are able to characterize the unique interior equilibrium of efforts.
- Empirical study of patent inventors
  - We use data from United States Patent and Trademark Office (USPTO).
  - We deal with the problem of endogenous matching between inventors and patents.
  - Results show a positive complementarity effect and a negative substitutability effect.
- Counterfactual analysis
  - We carry out a counterfactual study on the impact of innovation incentives on project's output.

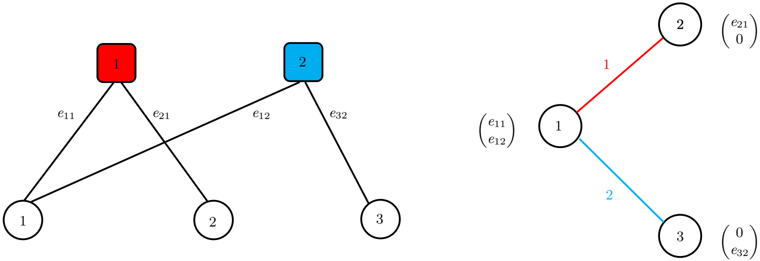
Production Function

- Assume there are $s \in \mathcal{P} = \{1, \dots, p\}$ projects (e.g., patents) and $i \in \mathcal{N} = \{1, \dots, n\}$ agents (e.g., inventors).
- The *production function* for project s is given by

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

- where y_s is the output of project s, e_{is} is the effort that agent i spent on project s ($e_{is} = 0$ if author i does not participate in project s),
- α_i captures the ability (fixed effect) of agent i,
- $g_{ij} \in (0, 1]$ measures the degree of compatibility between agents i and j,
- ϵ_s is i.i.d. project-specific random shock.
- the parameter λ represents the complementarity effect (spillover) between the efforts of collaborating agents, and
- $\mathcal{G}$ represents the *bipartite* network of agents and projects.

Example



1
2
e_{11}
e_{21}
e_{12}
e_{32}
1
2
3
$\begin{pmatrix} e_{21} \\ 0 \end{pmatrix}$
$\begin{pmatrix} e_{11} \\ e_{12} \end{pmatrix}$
$\begin{pmatrix} 0 \\ e_{32} \end{pmatrix}$

Utility

- The utility of agent i is given by

$$U_i(\mathcal{G}) = \underbrace{\sum_{s\in\mathcal{P}_i} \delta_s y_s}_{\text{payoff}} - \underbrace{\frac{1}{2}\left(\sum_{s\in\mathcal{P}_i} e_{is}^2 + \phi \sum_{s\in\mathcal{P}_i} \sum_{t\in\mathcal{P}_i\setminus\{s\}} e_{is}e_{it}\right)}_{\text{cost}}. \tag{2}$$

- where $\delta_s \in (0,1]$ is a discounting factor,[5] and
- The parameter ϕ measures substitutability between the efforts of the same agent in different projects.

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

Equilibrium Characterization

- Let

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

  where $\otimes$ denotes Kronecker product.
- $\mathbf{D}$ is an np-dimensional diagonal matrix given by $\mathrm{diag}_{s=1}^{p}\{\mathrm{diag}_{i=1}^{n}\{d_{is}\}\}$, where $d_{is} = 1$ if agent i is in project s and $d_{is} = 0$ otherwise.
- $\mathbf{G}$ is an $n \times n$ zero-diagonal matrix with the (i, j)-th $(i \neq j)$ element being g_{ij} (compatibility).
- $\mathbf{J}_p$ is an $p \times p$ zero-diagonal matrix with off-diagonal elements being ones.

Equilibrium Characterization

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

$$\rho_{\max}(\mathbf{L}^{\lambda,\phi}) < 1, \tag{4}$$

then the equilibrium effort portfolio is given by

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

where $\mathbf{L}^{\lambda,\phi} = \lambda\mathbf{W} - \phi\mathbf{M}$ represents the weight matrix of the line graph $\mathcal{L}(\mathcal{G})$, $\boldsymbol{\delta} = [\delta_1, \cdots, \delta_p]'$ and $\boldsymbol{\alpha} = [\alpha_1, \cdots, \alpha_n]'$, $\rho_{\max}(\mathbf{A})$ denote the spectral radius of a square matrix $\mathbf{A}$.

Data

- We use the dataset compiled by Bhaskarabhatla et al. (2021).
- The primary data consists of U.S. patents granted by USPTO, and then they match the assignees to publicly listed firms using Compustat data.
- Each entry in the dataset represents a patent-inventor pairing, containing information on both the patent and the associated inventor(s).
- For each patent, we know the application number, application year, and granted year. We merge the information of patent forward citation and patent value (Kogan et al. 2017) as measures of patent output.
- For inventors, we know their foreigner status, geographical location (at the state level for U.S.-based inventors), affiliated company, and industry classification, and past granted patents.

Data

To prepare the data for the estimation, we undertake the following sample selection procedure:

- We narrow our focus to patents filed in two specific industries: Semiconductors and related device manufacturing (with NAICS code 334413) and pharmaceutical preparation manufacturing (with NAICS code 325412).
- we focus on patents applied in 2003, in which the number of patent applications in the semiconductor industry is the highest compared to all other years.
- We exclude patents with missing information on the application year or associated inventors, and inventors with missing information on their characteristics.
- We drop inventors who only work on one solo-invented patent and the corresponding patents.

There are 8472 inventors and 6017 patents in NAICS 334413; 2888 inventors and 927 patents in NAICS 325412.

Go to Summary Statistics

Estimating the Production Function

- Following Equation (1), the production function of project s, with $s = 1, \ldots, p$, is given by

$$y_s = \sum_{i \in \mathcal{N}} \alpha_i d_{is} e_{is} + \frac{\lambda}{2} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} g_{ij} d_{is} d_{js} e_{is} e_{js} + \epsilon_s, \tag{6}$$

- We specify $\alpha_i = \exp(\mathbf{x}_i' \beta)$, where $\mathbf{x}_i$ is a $k \times 1$ vector of agent-specific exogenous characteristics.
- The empirical production function can be estimated by the nonlinear least squares (NLS) method or the maximum likelihood (ML) method (under the normality assumption on ϵ_s), with the unobservable e_{is} replaced by the equilibrium effort given in Equation (5).
- One potential problem of directly estimating Equation (6) is *endogeneity* of $\mathbf{D} = \mathrm{diag}_{s=1}^{p}\{\mathrm{diag}_{i=1}^{n}\{d_{is}\}\}$.

Endogenous Project Participation

- To address this endogeneity problem, we model the endogenous project participation of agent i in project s by

$$d_{is} = \mathbb{1}(z_{is}'\gamma + \xi\mu_i + \psi\eta_s + \kappa|\mu_i - \eta_s| + v_{is} > 0), \tag{7}$$

  where $\mathbf{z}_{is}$ is a $h \times 1$ vector of observables measuring compatibility between agent i and other agents participating in project s.
- The variable μ_i is an i.i.d.(0,1) agent-specific random component; and η_s is an i.i.d.(0,1) project-specific random component (Graham, 2016; 2017; Friel et al., 2016).
- v_{is} is a random component.

Endogenous Project Participation

- The production function (6) is then extended to

$$y_s = \sum_{i\in\mathcal{N}} \underbrace{\exp(\mathbf{x}_i'\beta + \zeta\mu_i)}_{\alpha_i} d_{is}e_{is} + \frac{\lambda}{2}\sum_{i\in\mathcal{N}}\sum_{j\in\mathcal{N}\backslash\{i\}} g_{ij}d_{is}d_{js}e_{is}e_{js} + \underbrace{\varsigma\eta_s + u_s}_{\epsilon_s}, \tag{8}$$

  to accommodate inventor and project specific unobservables, where u_s is assumed to be independent of η_s and normally distributed with zero mean and variance σ_u^2.
- Given $\mathbf{X} = [\mathbf{x}_i]$ and $\mathbf{Z} = [\mathbf{z}_{is}]$, the joint probability function of $\mathbf{Y} = (y_1, \cdots, y_p)$ and $\mathbf{D}$ can be specified as

$$\Pr(\mathbf{Y},\mathbf{D}|\mathbf{X},\mathbf{Z}) = \int_\mu\int_\eta \Pr(\mathbf{Y}|\mathbf{D},\mathbf{X},\mathbf{Z},\mu,\eta)\Pr(\mathbf{D}|\mathbf{Z},\mu,\eta)f(\mu)f(\eta)\,d\mu\,d\eta, \tag{9}$$

  from which we can estimate the parameter vector $\theta = (\lambda, \phi, \beta', \gamma', \xi, \kappa, \psi, \zeta, \varsigma, \sigma_v^2)'$.
- We estimate the model by Bayesian Markov Chain Monte Carlo (MCMC) sampling.

Estimation Results

Table: Estimation results for semiconductors and pharmaceuticals using patent forward citations.

		Semiconductors		Pharmaceuticals	
		(A) Exogenous Participation	(B) Endogenous Participation	(C) Exogenous Participation	(D) Endogenous Participation
Production					
Complementarity	(λ)	0.1538***	0.1677***	0.2782***	0.3192***
		(0.0233)	(0.0245)	(0.0663)	(0.0751)
Substitutability	(ϕ)	0.0485***	0.0502***	0.1663***	0.1871***
		(0.0062)	(0.0062)	(0.0204)	(0.0348)
Constant	(β_0)	-0.2110***	-0.2167***	-0.3956***	-0.4208***
		(0.0224)	(0.0243)	(0.0610)	(0.0670)
Foreigner	(β_1)	-0.0494***	-0.0499***	-0.0726**	-0.0711*
		(0.0139)	(0.0138)	(0.0365)	(0.0362)
Log accu. citations	(β_2)	0.1559***	0.1569***	0.2212***	0.2237***
		(0.0060)	(0.0062)	(0.0099)	(0.0106)
Seniority	(β_3)	-0.3823***	-0.3890***	-0.7042***	-0.7191***
		(0.0339)	(0.0327)	(0.1102)	(0.1120)
Inventor effect	(ζ)	–	0.0106***	–	0.0289*
			(0.0027)		(0.0169)
Patent effect	(ς)	–	0.0129***	–	0.0283
			(0.0027)		(0.0170)
Error term variance	(σ_ϵ^2)	0.8867***	–	0.7174***	–
		(0.0162)		(0.0338)	
Error term variance	(σ_u^2)	–	0.8664***	–	0.7170***
			(0.0162)		(0.0337)

[a] Model (A): assume exogenous participation of inventors on projects. Model (B): assume endogenous participation by Equation (7). The asterisks ***(**,*) indicates that its 99% (95%, 90%) highest posterior density range does not cover zero.

Estimation Results

Table: Estimation results for semiconductors and pharmaceuticals using patent forward citations.

		Semiconductors		Pharmaceuticals	
		(A) Exogenous Participation	(B) Endogenous Participation	(C) Exogenous Participation	(D) Endogenous Participation
Participation					
Constant	(γ_0)	–	-5.1908***	–	-5.0033***
			(0.0455)		(0.0884)
Location	(γ_1)	–	1.5446***	–	5.2940***
			(0.0674)		(0.2421)
Past coauthors	(γ_2)	–	8.4770***	–	7.2150***
			(0.1589)		(0.3919)
Common co-authors	(γ_3)	–	15.1374***	–	12.8617***
			(0.1659)		(0.3785)
Inventor effect	(ξ)	–	1.4049***	–	1.1192***
			(0.0647)		(0.0767)
Patent effect	(ψ)	–	-2.8966***	–	-2.6903***
			(0.0432)		(0.0960)
Homophily effect	(κ)	–	-3.6539***	–	-1.6792***
			(0.0669)		(0.0827)
Sample size					
Patents		6,017		927	
Inventors		8,472		2,888	

Conclusion

- we analyze the equilibrium efforts of agents who seek to maximize their utility when involved in multiple, possibly overlapping projects in a bipartite network.
- We show that both the complementarity effect between collaborating inventors and the substitutability effect between concurrent projects of the same inventor play an important role in determining the equilibrium effort level.
- We conduct a counterfactual analysis of the impact of the innovation incentive program on patent output. We find that the effectiveness of innovation incentives tends to be underestimated when the complementarity is ignored and overestimated when the substitutability is ignored.
- With some modifications, this framework can also be used to analyze competition between multi-product firms, formation of syndicated loans, and spillovers from science to innovations.

Counterfactual Analysis

- To illustrate the importance of accounting for the complementarity and substitutability effects in policy design and evaluation, we use our model to study a counterfactual incentive program to promote innovations.
- Under this program, we assume every inventor receives a reward, $r \in \mathbb{R}_+$, per unit of the output she generates. Then the utility function (2) of agent i can be extended to

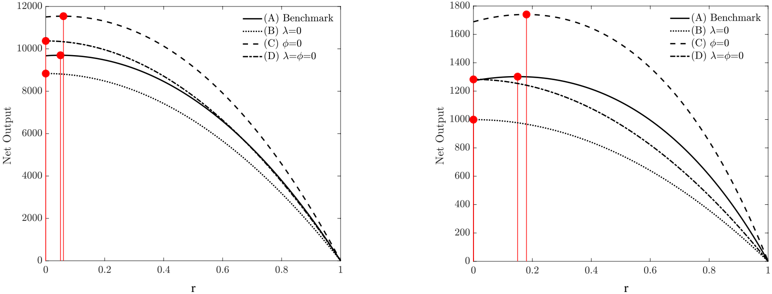
$$U_i(\mathcal{G}, r) = \sum_{s \in \mathcal{P}_i} (1 + r)\delta_s y_s - \frac{1}{2}\left(\sum_{s \in \mathcal{P}_i} e_{is}^2 + \phi \sum_{s \in \mathcal{P}_i} \sum_{t \in \mathcal{P}_i \setminus \{s\}} e_{is} e_{it}\right). \quad (10)$$

- Let $L(r) := L(r; \lambda, \phi) = \lambda(1 + r)W - \phi M$. We can show that, if $\rho_{\max}[L(r)] < 1$, then the equilibrium effort portfolio is given by

$$e^*(r) = (1 + r)[I_{np} - L(r)]^{-1} D(\delta \otimes \alpha). \quad (11)$$

- For a given reward rate r, the net total output can be computed as the total output, $\sum_{i \in \mathcal{N}} \sum_{s \in \mathcal{P}_i} \delta_s y_s(\mathcal{G}, r)$, minus the cost of the program, $\sum_{i \in \mathcal{N}} \sum_{s \in \mathcal{P}_i} r\delta_s y_s(\mathcal{G}, r)$.

Counterfactual Analysis



((A)) Semiconductors ((B)) Pharmaceuticals

Figure: The net total output in the presence of a merit-based reward, $\sum_{i\in\mathcal{N}}\sum_{s\in\mathcal{P}_i}(1-r)\delta_s y_s(\mathcal{G},r)$, for the semiconductor industry and the pharmaceutical industry. The maximum at the optimal rate, r^*, is highlighted with vertical lines for different model specifications.

Data

Table: Sample statistics for semiconductors.

	Min	Max	Mean	S.D.	Obs.
Patents					
Forward Citations (in log)	0.0000	5.7398	1.6047	1.1063	6017
Value (million dollars in 1982 prices)	0.0209	162.0552	7.4947	9.9904	6017
Number of inventors (in each patent)	1	13	2.5784	1.4803	6017
Inventors					
Foreigner	0	1	0.4214	0.4938	8472
Log life-time patent citations	0	9.3284	2.7749	1.7108	8472
Log life-time patent values (millions)	0	6.4146	1.9198	1.0437	8472
Seniority (decades since first granted patent)	0	2.5000	0.2275	0.4088	8472
Number of patents (for each inventor)	1	49	1.8312	1.9784	8472

Notes: Semiconductor and related device manufacturing (NAICS: 334413). This sample is based on patents applied in 2003. We drop inventors with a single solo-invented patent and the corresponding patents. We add one before taking the log of forward citations.

Data

Table: Sample statistics for pharmaceuticals.

	Min	Max	Mean	S.D.	Obs.
Patents					
Forward Citations (in log)	0.0000	5.1761	1.3639	1.2304	927
Value (million dollars in 1982 prices)	0.0100	563.4050	51.2777	60.5520	927
Number of inventors (in each patent)	1	24	4.2945	2.9925	927
Inventors					
Foreigner	0	1	0.2753	0.4467	2888
Log life-time patent citations	0	8.3156	2.4375	1.7482	2888
Log life-time patent values (millions)	0	6.8764	3.4198	1.2467	2888
Seniority (decades since first granted patent)	0	2.5000	0.2617	0.4495	2888
Number of patents (for each inventor)	1	18	1.3785	0.9459	2888

Notes: Pharmaceutical preparation manufacturing (NAICS: 325412). This sample is based on patents applied in 2003. We drop inventors with a single solo-invented patent and the corresponding patents. We add one before taking the log of forward citations.

Return