The productivity of science & engineering PhD students hired from supervisors' networks

Stefano Baruffaldi [a], Fabiana Visentin [a,b,*], Annamaria Conti [c]

[a] Chair of Economics and Management of Innovation, École Polytechnique Fédérale de Lausanne, ODY 1.18 Station 5, Lausanne 1015, Switzerland
[b] BRICK, Collegio Carlo Alberto, via Real Collegio 30, 10024 Moncalieri, Italy
[c] Scheller College of Business and RCEA, Georgia Institute of Technology, 800 West Peachtree Street, Atlanta, GA 30308, USA

ARTICLE INFO

ABSTRACT

We compare the scientific productivity of PhD students who are hired from a fine-grained set of mutually exclusive affiliation types: a PhD supervisor's affiliation, an external affiliation from which the supervisor derives her coauthors, and an external affiliation with which the supervisor has no coauthorship ties. Using a novel dataset of science and engineering PhD students who graduated from two major Swiss universities, we find that the most productive PhD category is the one made of students who are affiliated with universities other than their supervisors' affiliation, but from which the PhD supervisors derive their coauthors. This result suggests an inverted U-shaped relationship between PhD students' productivity and the social distance from their supervisors. Additionally, we find evidence consistent with the role of supervisors' coauthor networks in resolving information asymmetries regarding PhD talent.

© 2015 Elsevier B.V. All rights reserved.

1. Introduction

To date, academia remains a major locus of knowledge production, generating a considerable fraction of research articles (National Science Board, 2008). Within academia, knowledge production largely occurs in laboratories, which are increasingly organized as small family businesses run by their laboratory heads (Freeman et al., 2001). In these laboratories, there is no doubt that PhD students play a fundamental role. A brief glance at their supervisors' websites convinces one that PhD students represent an important proportion of academic research groups and produce a large share of the groups' scientific articles (Conti and Liu, 2015a, 2015b; Conti et al., 2014; Stephan, 2012). Despite their acknowledged contributions[1], few studies have attempted to analyze the determinants of PhD student productivity. Such works have primarily focused on the impact of faculty quality on PhD student outcomes and on the difference in productivity of foreign versus domestic students (Black and Stephan, 2010; Waldinger, 2010; Gaule and Piacentini, 2013).

We contribute to the existing literature on PhD students by relating their productivity to different categories of affiliations from which they obtained their master's degree. Specifically, we distinguish between three sets of mutually exclusive affiliation types: a PhD supervisor's affiliation, an external affiliation from which the supervisor derives her coauthors, and an external affiliation with which the supervisor has no coauthorship ties. The rationale for making this distinction is that supervisors face important information asymmetries when hiring PhD applicants and a way to overcome them is to hire applicants from affiliations belonging to their network, *selection effect*. Upon the hiring of their PhD students, supervisors can further use membership in their network as a criterion to allocate, *ex-post*, their limited resources among the students, *treatment effect*.

For our analysis, we use a novel dataset of 4666 PhD students in science and engineering who graduated from two major Swiss technology institutes: the Swiss Institute of Technology of Lausanne (EPFL) and the Swiss Institute of Technology of Zurich (ETHZ). For these students we collected information about their master's affiliations as well as information about their biographical characteristics and publication records. We matched these data with fine-grained information regarding the PhD students' supervisor characteristics, including their publications, patents, and the size

[1] For PhD students' contributions outside of academia, refer to Conti and Visentin (2015a, 2015b) and Sauermann and Roach (2012).

of their coauthor network. This hand-collected dataset allows us to be precise about the link between the PhD students' research output and their master's affiliation type.

We show that the most productive PhD students are the ones who are affiliated with universities other than their supervisors' affiliation, but from which the PhD supervisors derive their coauthors. This finding indicates that while supervisor connections are important to either overcome information asymmetries or allocate resources *ex-post* among PhD students, when these connections are too strong, they may become detrimental to the students' productivity. Hence, the resulting functional form between PhD students' productivity and the social distance from their supervisor is an inverted U-shape. We further explore our initial result in an attempt to understand the reason why PhD students of a university from which a supervisor derives her coauthors are more productive than students from an affiliation with which the supervisor has no coauthorship ties. In particular, we aim to assess whether our result is primarily driven by a selection or a treatment effect. To achieve this goal, we examine two instances in which membership in a research network leads to different outcomes, depending on whether the selection or the treatment are the main drivers of our results. The first instance we consider is one in which a student has published prior to beginning her PhD. If one student has a record of pre-PhD publications that a professor can use to assess the student's quality and we observe that there is a substitute relationship between these publications and being affiliated with a university from which the supervisor derives her coauthors, we can reasonably expect that one information source is replacing the other one in the determination of a student's research aptitude. The second instance we consider is one in which a supervisor has intense coauthorship ties with a student's master's university. If the selection mechanism prevails, having only one coauthor from a given university should be sufficient to attenuate information asymmetries, while stronger connections should play no role (Granovetter, 1973). The results of these analyses are consistent in suggesting that selection is an important motive for choosing students from universities from which the students' supervisors derive their coauthors.

Our results speak to the existing literature on the role of networks in academia (Horta et al., 2010; Zinovyeva and Bagues, 2015; Li, 2012; Laband and Piette, 1994; Brogaard et al., 2014; Baruffaldi and Landoni, 2012)[2]. These studies have analyzed the impact of in-network status on promotions (Zinovyeva and Bagues, 2015), grant assignments (Li, 2012), and publications in scientific journals (Laband and Piette, 1994; Brogaard et al., 2014). The paper closest to ours is the one by Horta et al. (2010) wherein they closely examine the hiring of professors from those universities that trained them, also known as academic inbreeding. In contrast, with this paper we refer to an early stage of the academic career, the students' advancement from master's studies to PhD studies. We provide a fine-grained taxonomy of individuals staying in the same university, distinguishing within the category those students who obtained their master's outside of their supervisors' affiliation from those who were and were not trained at an affiliation from which the supervisors derive their coauthors. Furthermore, we propose a descriptive analysis in an effort to disentangle selection from treatment effects.

The rest of this paper is organized as follows: Section 2 presents the conceptual framework, Section 3 describes the empirical context, Section 4 discusses the main variables and the empirical method, Section 5 presents the empirical findings, Section 6 discusses a robustness check that further validates our findings,

Section 7 provides results regarding the distinction of the selection from the treatment effect, and Section 8 concludes.

2. Conceptual framework

A good PhD candidate needs to possess a multiplicity of characteristics such as an aptitude for research, technical skills, motivation, a capacity for conducting independent work, and fit within the supervisor's group (Stephan, 2012). Some of these characteristics are identifiable, at relatively low cost, from the students' curricula. Other characteristics, the so called 'unobservable habits of action', cannot be easily gauged and would require a supervisor to rely on indirect assessment procedures (Arrow, 1972). One such procedure consists of selecting students from an affiliation with which the supervisor is connected. In this way, the supervisor can gather information on the applicants' characteristics by using her knowledge about these affiliations or by relying on the members' referrals (Montgomery, 1991). Conditional on the hiring of a group of PhD students, a supervisor can use the students' affiliation with one of her connected universities as a criterion for allocating, *ex-post*, her limited resources among admitted PhD applicants. Referring to the large amount of literature on homophily (for a survey of this literature refer to McPherson et al., 2001), we should expect larger resources to be allocated to students that are socially close to their supervisor. The use of ties to overcome, *ex-ante*, asymmetric information problems and to allocate, *ex-post*, supervisors' limited resources suggests that PhD students who were trained at affiliations with which their supervisors have ties may be more productive than students who were not.

Despite what we just argued, there are arguments to be made against populating research groups with PhD students who come from affiliations with which a supervisor has developed connections. In fact, it is possible that these students are hired not because they are productive, but because their supervisor is doing a favor for one of her ties (Prendergast and Topel, 1996; Durante et al., 2011) or is trying to save on search costs (Greif, 1993). It is plausible to expect that the loss in terms of PhD student productivity is larger, the stronger the ties are between a supervisor and the affiliations from which the supervisor derives her PhD students. This point can be clearly illustrated by making a comparison with the way we structure the relationships with our friends. Typically, it is more difficult to refuse a favor to a close friend than to a not-so-close one, implying that the likelihood that somebody indulges in favoritism is higher with the first type of friend that with the second. Additionally, because we trust close friends, when we receive a recommendation from them, we are not as thorough as we would be with not-so-close friends in verifying the validity of the recommendation. In other words, we are willing to save on verification costs, even though the recommendation may not ultimately maximize our benefit. Based on these arguments, our hypothesis is that there is an inverted U-shaped relationship between a PhD student's social distance from her supervisor and the student's research productivity. Thus, populating a research group with students coming from affiliations with which a supervisor has developed ties increases the productivity of the students, but when the ties become too strong the costs prevail over the benefits and the students' productivity decreases.

3. Context

We examine a sample of PhD students from EPFL in Lausanne and from ETHZ in Zurich. These universities are the two Federal Institutes of Technology in Switzerland. EPFL is located in the French-speaking region of Switzerland while ETHZ is in the German-speaking region. EPFL and ETHZ are responsible for a large portion of the research in science and engineering that is produced

---

[2] For a more general discussion of the relevance of networks, refer to Uzzi and Lancaster (2003) and Uzzi and Gillespie (2002).