Available online at www.sciencedirect.com

ScienceDirect

Journal of Economic Theory 147 (2012) 713–737

JOURNAL OF Economic Theory

www.elsevier.com/locate/jet

Random walk to innovation: Why productivity follows a power law ☆

Christian Ghiglino *

Department of Economics, University of Essex, Wivenhoe Park, Colchester, CO4 3SQ, UK

Received 16 July 2009; final version received 10 September 2010; accepted 16 February 2011

Available online 8 March 2011

Abstract

This paper examines a search model in which knowledge is created as rational, but poorly informed, innovators obtain new ideas from combinations of existing ideas. We assume that the productivity of an idea is stochastic and depends on the productivity of the parental ideas. Importantly, we assume that the contribution of these parents to the productivity of the final idea is enhanced by prior use of these in knowledge creation. We identify conditions on the search costs leading to two properties: 1) the tail of the distribution of the productivity of innovations is a power law, and 2) the number of citations, i.e., times an idea is used in the process of innovation, follows a displaced power law. Both these properties are consistent with the available empirical evidence on the productivity of innovations and on patent citations.

JEL classification: O31; O41

Keywords: Economic growth; Technological progress; Innovations; Random growing networks; Ideas; Scale-free distributions; Patent citations

1. Introduction

The view that innovation occurs as existing ideas are recombined in new ways can be traced back to the early developments of epistemology, and in economics back to at least Schumpeter.

☆ We thank A. d'Autumne, R. Bailey, R. Becker, M. Boldrin, B. Decreuse, S. Goyal, T. Hens, M. Jackson, C. Jones, G. Kemp, N. Kuschy, C. Le Van, T. Mitra, P. Peretto, B. Rogers, K. Shell, J. Stachurski, P. Young and J. Zweimuller for useful comments. We are particularly grateful to three anonymous referees and an Associate Editor of this Journal for very useful remarks and suggestions.

* Fax: +44 1206 872724.
E-mail address: cghig@essex.ac.uk.

0022-0531/$ – see front matter
doi:10.1016/j.jet.2011.02.004

Recently, Weitzman [32][1] proposes a model in which "knowledge can build upon itself in a combinatorics feedback process". However, this feedback is not exploited and progress is driven by the growth in the stock of available ideas.[2] The quality of ideas does play a role in search-based models of innovation. For example, in Kortum [22] researchers find new ideas via blind draws from a given distribution of productivity; and the shape of the distribution matters. However, the distribution of productivity of ideas is not itself related to the primitives of the economy and is not affected by the behavior of present and past innovators. As noticed in Luttmer [25], this weakness is shared by most of the literature on innovation, and on firm and city dynamics.

The aim of the paper is to construct a plausible mechanism by which rational innovators obtain new ideas, resulting in productivity of ideas distributed according to a stationary distribution with a power law upper tail. We assume that in the first step of innovation ideas are produced as combinations of existing ideas. Naturally, the quality of the parental ideas is an important factor in the productivity of the final idea. However, this process of "reproduction with heredity" is not sufficient to generate a distribution of productivity with a fat tail. A further element is needed. We assume that the contribution of the parents to the productivity of the final idea is improved by prior use in knowledge creation. We call this process "learning-by-use".

As an example, consider the method for making a self-lubricating liner as described in US Patent 4 419 164. This idea has been used to produce several ideas related to the make of different types of magnetic discs: rigid, flexible, encapsulated, etc. We believe that the widespread use of the idea "self-lubricating liner" in innovation has increased its potential contribution to the productivity of any new idea to produce a final good. Note that, "learning-by-use" affects only the productivity of ideas when used as inputs in knowledge creation.

The general environment is common to search-based models of innovation, for example to Kortum [22] which itself adopts the quality-ladder structure of Grossman and Helpman [17]. There is a non-expanding, countable, but large number of final goods and only one input, labor. Each final good is produced in a sector characterized by a state-of-the-art technology, called an "idea", and its associated productivity (a scalar). Infinitely lived agents are either innovators, who work in an R&D sector that produces new ideas, or workers, in which case they work in a firm producing a final good.

Innovation within the R&D sector, in the form of higher productivity in final good production, takes place in two steps. First, new ideas are produced by innovators using existing ideas, a process taking place in a short lapse of time. Second, each innovator develops his idea to the stage of a prototype, its productivity being evaluated and compared to the state-of-the-art. While the idea is patented only if it improves an existing patent, all prototype ideas are added to the set of existing ideas for further use in knowledge creation.

We consider an environment in which the innovator cannot, at the level of the first step of innovation, selectively choose input ideas from the stock of ideas to create new ideas.[3] We assume semi-blind search in the sense that the innovator can proceed to blind draws from the pool of existing ideas and that each draw reveals the list of parents, or references, of the drawn idea but not its productivity. This type of search is analogous to a tourist exploring a city by walking along its streets without a map. In the model, the innovator explores the network of linkages across ideas,

[1] See also Auerswald, Kauffman, Lobo and Shell [5]. The empirical justification for this approach can be found in the above reference as well as in Weitzman [32].

[2] As is also the case in idea-based endogenous growth models.

[3] The impossibility of selection is common to search models of innovation, e.g., "the blind draws" in Kortum [22].

for example, using the information in a patents' database, or the bibliographical references in scientific papers.

The model generates endogenously a stationary distribution of productivity of ideas, which is shown to be, at least asymptotically, a power law. Note that a power law upper tail implies that the largest order statistic is a Frechet distribution with a fat tail, a property that would not arise with, for example, a log normal distribution. This is relevant because state-of-the-art ideas, and therefore patents, are expected to follow an extreme value distribution with properties that agree with the available empirical evidence.[4] A second important empirical stylized fact concerns the use of prior knowledge in patent behavior: the number of citations received by patents follows, asymptotically, a power law. The prediction of the present theoretical model is also consistent with this feature.

The intuition underlying the predictions of the model is as follows. The productivity of an idea to produce a final good depends on the productivity of its parents. Recursively, the productivity of parents depends not only on the productivity of the grandparents but also on how many times the parents of the new idea have been used in knowledge creation. The semi-blind search of the innovator is then oriented towards ideas that have been often used in knowledge creation. How does the innovator find them? By selecting the ideas among the parents of randomly drawn ideas, that is, the "friends of randomly met friends". It is known (e.g., Vazquez [31]) that such behavior mimics linear preferential attachment in the sense of Barabasi and Albert [6]. The endogenously formed network of linkages is in this case characterized by a distribution of in-degrees (i.e., how many times an idea is used) that is asymptotically a power law. Finally, as productivity of ideas depends, ceteris-paribus, on the in-degree of the parents, this implies that productivity also follows a power law.

In the analysis we are able to link the existence and properties of the stable distribution (under innovators' behavior) of the productivity of ideas to the linkages across parental ideas. The endogenous network of links is then related to the search strategies. In particular, the conditions for a power law upper tail in the distribution and the value of its exponent, are related to the proportion of innovators choosing a network search vs. a pure random search. Finally, this choice is related to the individual cost of search as compared with the expected return from innovation. Note that the analysis is related to Jackson and Rogers [18] and Vazquez [31] who also look at environments with a mix of agents adopting network searches and agents adopting random searches. However, while in this literature the agents are "born" with a strategy, we link the endogenous network formation mechanism to the existence of spillovers across ideas (making search profitable), and to heterogeneity in search costs.

1.1. Related literature

The present paper is related to several other strand of the literature.

Besides being inspired by Weitzman [32], the present paper is related to Auerswald, Kauffman, Lobo and Shell [5] who assume that trials by firms reveal the topology of a pre-existing distribution of uncovered recipes (ideas). Progress is incremental in the sense that close recipes (according to a well-defined notion of distance) have more chances of being uncovered. Most importantly, in Auerswald et al. [5] externalities arise between these uncovered close recipes. The similarity with our model lies in that new (uncovered) ideas benefit from existing ideas and

[4] See the discussion in Jones [19]. Note also that many theoretical models of innovation assume power law tails in order to reproduce the stylized facts of growth.

also that the productivity of existing ideas improves progressively as neighbor recipes are discovered. Differently from this paper, learning is mechanical, via blind trials, without any optimizing behavior, and the technological landscape is exogenous.

This paper is also related to the literature that applies general equilibrium models to obtain firm size distributions, for example Luttmer [25]. A common feature of these models is that in a world in which the productivity of incumbent firms increases over time, stationarity in the size distribution requires sustained new entry, which itself requires that entrants take advantage of the improvement of the incumbents. Also in this vein, Luttmer [26] in this issue, considers the case of new firms that benefit, after entering the market, from the successes of incumbent firms through stochastic shocks, and highlights a positive role of trial and error for growth.[5] The hypothesis that both incumbents and entrants innovate is modeled within a Schumpeterian quality-ladder model in Acemoglu and Cao [1]. Overall, this literature shares with this paper the idea that continuous inventive activity by incumbents is beneficial for the entrant innovator.

The cited literature on firm sizes is a subset of a vast recent, and less recent, literature on power laws in economics and elsewhere, for example, Gabaix [15] and Cordoba [9] for city sizes and Benhabib et al. [7] for the wealth distribution (see Gabaix [16] for a survey). A "universal" way to generate power laws is to assume proportional growth with new entry, a notion proposed initially by Yule [33], formalized by Simon [29] and, for the case of random growing graphs, revisited as "linear preferential attachment" by Barabasi and Albert [6]. Although the mechanics of how entrants link to existing nodes is the common feature, these models are disparate in the reasons why this happens. In particular, in Jackson and Rogers [18] apparent preferential attachment is a by-product of the type of search followed by agents. However, agents' choice of the type of search is not endogenized, and the mechanical aspect of the model reappears. In the present model the productivity of ideas does increase linearly with prior use, but this does not translate mechanically into proportional growth. "Learning-by-use" gives the "right" incentives to innovators to adopt specific search strategies, and indirectly, by making agents mimic preferential attachment, leads to proportional growth.

Related to this paper, an older strand of the quality-ladder literature considers sequential innovations in a partially ordered network of commodities. Kelly [21] proposes a model in which innovation in sector j is affected by research in sector $j-1$ and sector $j+1$. However, the network of spillovers, which can be viewed as a special case of a predetermined technological landscape, is exogenously given.

Finally, the presence of linkages across sectors is at the core of a recent literature using input–output models to study volatility in productivity (see in particular, Acemoglu, Ozdaglar and Tahbaz-Salehi [2] and Carvalho [8] and the papers cited there). These models revisit the classical input–output formalism and consider the associated adjacency matrix and directed graph. They show how the propagation of an idiosyncratic shock through the economy is related to the position of the shocked sector in the graph. Then the properties of the equilibrium path are related to the properties of the adjacency matrix. The similarity with our model is that higher order connections (as the parents of parents) matter for productivity.

The paper has the following structure. In Section 2 the underlying economic model is described. In Section 3 the search environment and the main preliminary results are obtained. In Section 4, the results of Section 3 are used to derive the stable stationary equilibrium distribution of the productivity of ideas. Section 5 concludes, while Appendix A contains the proofs.

[5] This is related to Alchian [4], Nelson and Winter [28] and Jovanovic [20].

2. The model

2.1. The environment

There is a countable set of dates, $t \in \{1, 2, 3, \ldots\}$, but the horizon is infinite. Consumption goods come in a given, countable and large but finite number q of varieties indexed by s. These are produced using labor and ideas. In each sector, the technology to produce the variety is linear, $y_{st} = \hat{a}_{it} L_{st}$, where y_{st} is output, $\hat{a}_{it}$ is the productivity of an idea i used to produce output s and L_{st} is labor input. Consumption good s is produced by a monopolist that holds the patent for the state-of-the-art idea in sector s and charges a price p_{st} that depends also on the productivity of the second best idea. Ideas are produced in the R&D sector using specific labor and existing ideas. There is no capital.

At time t there is a countable and large but finite number L_t of infinitely lived agents and population is growing at a constant rate g. Agents are born either as innovators developing new ideas, or as workers supplying at time t one unit of time to produce one of the final consumption goods for a wage w_t. In agreement with the stylized facts we assume that the total number of individuals engaged in research at time t, R_t, grows at the rate g.[6] Agents discount the future at a rate ρ and have symmetric standard preferences over all goods. Agents are assumed to be price takers. The return from innovation is random. However, as agents are risk neutral and can freely borrow at an interest rate r_t, maximizing utility over consumption implies that innovators maximize the present value of net expected returns from innovation.

2.2. The R&D sector

As is usual in search-based growth models, the innovator is assigned randomly to a sector in which to innovate. We assume that new ideas are produced by each innovator at a unit rate. An innovation, in the form of higher $\hat{a}_{it}$, takes place in two steps. First, innovators obtain new ideas by combing existing ideas. Formally, when a new idea i is created it draws on existing ideas, called the *parents* of an idea i. For simplicity we assume that any idea has m, $m > 1$, parents, typically designated by a j.[7] The details of this instantaneous search process for existing ideas are described in the next section.

In the second step, the idea is developed as a prototype and its productivity evaluated. If the efficiency of an idea i, $\hat{a}_{it}$, exceeds the state-of-the-art in the associated sector, the idea is patentable and itself becomes the state-of-the-art. In this case the patent produces a stream of returns until it is surpassed. All prototype ideas are also added to the set of existing ideas for further use in knowledge creation.

2.3. Productivity of ideas

There are two notions of productivity depending on whether the idea is used to produce a final good or as an input in knowledge creation. The productivity of an idea i to produce the corresponding final good at time t, $\hat{a}_{it}$, has a stochastic component but also depends on the quality of the parental ideas. The quality of each parent j, a_{jt}, depends on its final good productivity $\hat{a}_{jt}$

[6] Alternatively, the research effort R_t could be endogeneized along the lines of Kortum [22].

[7] Similarly j has as offspring an idea i if an idea j enters in the combination composing an idea i.

but also on its applicability, which depends on its prior use in knowledge creation. The economic interpretation of this effect, denoted "learning-by-use" is discussed in the Introduction. Note that a_{jt} is then the productivity of an idea j when used in knowledge creation but evaluated in terms of its contribution to the new final idea to produce the final consumption good.

We assume that the productivity $\hat{a}_{it}$ of an idea to produce a final good is the productivity of the recipe at the time it is discovered, that is, for an idea i obtained at time $t_i \leqslant t$, $\hat{a}_{it} = a_{it_i}$. This implies that the productivity in producing the final good is not improved by further use of its parent ideas, while the productivity of parent ideas when used as inputs in the production of ideas do improve with their use.

We now state the formal assumption concerning productivity of ideas. Let N_t be the stock of ideas at time t. Let $\mathcal{N}(i,t)$ (with $Card(\mathcal{N}(i,t)) \geqslant m$) be the set of parents and offspring of an idea i at time t. Let $\widetilde{A}_{it}$ be the idiosyncratic component of the productivity of an idea i and assume that the components of $(\widetilde{A}_{it})_{i=1}^{N_t}$ are independently extracted from a known and fixed distribution, uncorrelated with the parents' productivity. The formal definition of productivity is as follows.

Assumption 1. The total productivity[8] a_{it} of an idea i when used in knowledge creation at time t is given by

$$\begin{aligned} a_{it} &= \widetilde{A}_{it} + b_P \sum_{j \in \mathcal{N}_P(i)} \theta_{ijt} a_{jt} + b_O \sum_{h \in \mathcal{N}_O(i,t)} \theta_{iht} a_{ht} \\ &= \widetilde{A}_{it} + b_P \sum_{j \in \mathcal{N}(i,t)} g_{ijt} \theta_{ijt} a_{jt} \end{aligned}$$

where b_P and b_O are positive scalars, $\theta_{i1t}, \ldots, \theta_{i\mathcal{N}(i,t)}$ are independent random variables following a distribution $N(1, \sigma_\theta)$ with $\sigma_\theta \ll 1$, $\widetilde{A}_{it} \sim N(\overline{A}, \sigma)$ with $\sigma \ll \overline{A}$. Furthermore, $\mathcal{N}_P(i)$ is the set of parents of an idea i, $\mathcal{N}_O(i,t)$ is the set of offspring of an idea i, and the $\theta_{ijt} g_{ijt}$ represent the sectorial links, where g_{ijt} is either 0 or 1 depending on whether the link exists or not.

The Hadamard product $G_t \cdot \theta_t$, noted as $G_t(\theta_t)$, and such that $G_t(\theta_t)_{ij} = \theta_{ijt} g_{ijt}$ allows reformulation of Assumption 1 as

$$A_t = \widetilde{A}_t + b_P G_t(\theta_t) A_t$$

where A_t and $\widetilde{A}_t$ are vectors of dimension N_t, with components a_{it} and $\widetilde{A}_{it}$ respectively. Provided the inverse exists, we obtain

$$A_t = \big(I - b_P G_t(\theta_t)\big)^{-1} \widetilde{A}_t$$

To ensure existence of the inverse, we choose b_P smaller than the absolute value of the reciprocal of the largest eigenvalue of the modified adjacency matrix $G_t(\theta_t)$. The following expansion is then convergent and well defined

$$A_t = \sum_{\tau=0}^{\infty} \big(b_P G_t(\theta_t)\big)^{\tau} \widetilde{A}_t$$

[8] Negative productivities are not ruled out in the formalism but will play no role in our economic analysis, in particular because of the asymptotic nature of the results of Theorem 1. Note also that each random component in a_{it} is drawn exactly once.

From this expression we can see that, besides stochastic independent elements, productivity at time t depends on the linkages across all ideas, summarized by the number of offspring of each idea, that is, the "in-degree" of each idea.

Remark. Any of the neighbors of an idea i at time t may contribute to the productivity of the idea, independently of the sector for which it was initially designed. The assumption that there is only one network of ideas implies the existence of spillover across recipes producing different goods.

2.4. Definition of sub-equilibrium and equilibrium

In the present environment, the state of the economy at time t can be defined in terms of two sets of variables. First, let X_t denote the list of variables describing the R&D activity, that is, the productivity of the N_t ideas in both final good and knowledge production, $(\hat{a}_{it}, a_{it})_{i\in N_t}$, the network of linkages representable by a matrix, $\mathcal{N}_t$, and the set of productivities of patents active at time t, $(\chi_{st})_{s=1}^{q}$. Second, let E_t be the list of variables describing the general activity of the economy, that is, the population L_t, the consumption prices p_{st} and the interest rate r_t. We consider the wage w_t as the numeraire, $w_t = 1$ all t. As there is uncertainty regarding the success of innovations all the values are stochastic and the variables are, in principle, random processes.

We consider two notions of equilibrium, a sub-equilibrium and an equilibrium path. In a sub-equilibrium the research activity of the R&D sector is not endogeneized, i.e., the behavior of the innovators is not modeled. Equivalently, in the following definition, it is assumed that the behavior of the innovators is irrelevant.

Definition 1.

1. A **sub-equilibrium** is a sequence $(X_t, E_t)_{t\in\mathbb{N}}$, such that for all $t \geqslant 0$, at time t
   (a) firms maximize profits and agents maximize their expected utility for given state E_t and using strategies depending only on the realization of variables at date t;
   (b) the sequence $(X_t)_{t\in\mathbb{N}}$ is such that ideas of productivity $(\hat{a}_{it})_{i\in N_t}$ are obtained by random draws from a given distribution Ξ, independently of $(a_{it})_{i\in N_t}$ and of the network of linkages $\mathcal{N}_t$;
   (c) all markets clear.
2. An **equilibrium path** is a sub-equilibrium $(X_t, E_t)_{t\in\mathbb{N}}$ such that, for all $t \geqslant 0$, $(\hat{a}_{it}, a_{it})_{i\in N_t}$ and $\mathcal{N}_t$ are generated by a dynamic process in which all innovators optimize using strategies depending only on the realization of variables at date t.

Kortum [22] considers a similar economy in which the activity of the R&D sector boils down to simple random draws from a Pareto productivity distribution, and goes on to derive the general equilibrium path. We instead consider a reduced form of the model that focuses on the detailed activity of the R&D sector. This motivates the following assumption concerning the existence of a sub-equilibrium, while a proof of the existence of an equilibrium path, which is the aim of the paper, appears in Section 4.

Assumption 2. Let Ξ represent a distribution with unbounded support and a power law upper tail such that, for large k,

$$P[x > k] \simeq k^{-\xi}$$

where $\xi > 1$. Assume that the ideas obtained by the innovators are such that productivity $\hat{a}_{it}$ is a random draw from a distribution Ξ with a power law upper tail of parameter ξ. Then there exists a sub-equilibrium $(X_t, E_t)_{t\in\mathbb{N}}$ with productivities $(\hat{a}_{it})_{i\in N_t}$

Finally, let the present value at time t of the expected return from innovation, in a sector s, be a function V_s of the state of the economy $V_s(\hat{a}_{it}, \widehat{X}_{it}, E_t)$ where $\hat{a}_{it}$ is the productivity of an idea i to produce a final good s, and $\widehat{X}_{it}$ is as X_t but without the productivity of an idea i, $\hat{a}_{it}$. We assume that, given the state of the economy $(\widehat{X}_{it}, E_t)$ at time t, and for any sub-equilibrium path $(\hat{a}_{it}, \widehat{X}_{is}, E_s)_{s=t}^{\infty}$, V_s is increasing in the productivity $\hat{a}_{it}$.

3. Behavior of rational innovators

In this section we begin by describing the search environment and the set of strategies available to the innovator and derive the evolution of the network $\mathcal{N}_t$ of linkages across ideas as a function of the strategy chosen by the innovators (Lemma 1). In Lemma 2 we characterize the strategy that is individually rational when the innovator faces a specific, exogenously given, network of linkages and a given productivity distribution.

3.1. Search costs and set of strategies

We assume semi-blind search in the sense that the innovator makes blind draws from the pool of existing ideas, each draw revealing the list of parents, or references, of the drawn idea j but not its productivity a_{jt}. The agent then faces a sequence of decisions. After each draw j, the agent has to decide whether to stop the search, to make a further draw from the entire pool of ideas or to select at random a parental link of the drawn idea j. The choice depends of course on the cost of search. We assume that at time t a given cost, expressed in terms of the wage (taken as numeraire), is incurred by each innovator searching for existing ideas. This cost depends on the length of search, or equivalently, on the number of draws. We assume also that search is anonymous, memoryless and does not directly reveal the productivity of ideas.

Assumption 3. An innovator h searches for ideas to be combined subject to a cost $c_{ht}\tau$ that is linear in the length of the search τ (the number of random draws), with given $c_{ht} > 0$. Importantly, the list of references, or parents, but not the productivity of ideas is revealed during the search. Finally, we assume that the search is anonymous and memoryless.

To characterize formally the innovator's behavior we need to characterize his set of strategies, i.e., the sequences of his decisions. Two mild preliminary assumptions are: 1) any search starts with a random draw from the set of existing ideas, and 2) strategies are sequences of finite length. While the first part is trivial the second part is implied by the fact that returns from innovation are finite. Finally, we assume that searching for ideas happen more quickly than development of the prototype, i.e., search happens instantaneously.

Definition 2. At time t and for an innovator h, a strategy to find an idea is a finite sequence $(p_{h,t,0}, q_{h,t,0}, p_{h,t,1}, q_{h,t,1}, p_{h,t,2}, q_{h,t,2}, \dots, p_{h,t,T}) \in \{0,1\}^{2T+1}$ with $T < \infty$, $p_{h,t,0} = 0$, $q_{h,t,0} = 0$ and $p_{h,t,T} = 1$ where

1. $p_{h,t,\tau} = 1$ represents the decision to stop searching after draw τ and $p_{h,t,\tau} = 0$ the decision to continue searching after draw τ.
2. $q_{h,t,\tau} = 1$ represents the decision, conditional on $p_{h,t,\tau} = 0$, to execute a random draw $\tau + 1$ from the list of references of the idea obtained at draw τ and $q_{h,t,\tau} = 0$ the decision to execute a random draw from the pool of ideas.

All the strategies in which the search ends by selecting a node at random from the entire population give the same expected productivity, independently of the sequence of draws leading to that final draw. Therefore, out of the set of such strategies only the strategy with a unique random draw from the total stock of ideas is relevant. This strategy is called **Strategy** I or **random search**. The complementary strategies are such that the idea is randomly selected from the list of parents of an idea previously selected. Again, only strategies involving a single initial draw from the stock of ideas are to be considered. These strategies using **network search** are called **Strategy** II-l, where l is the length of the strategy sequence, or "walk", after the initial random draw from the stock.

Definition 3. We consider the following strategies:

1. **Random search**, **Strategy** I. The innovator chooses randomly an idea: $p_{h,t,1} = 1$.
2. **Network search**, **Strategy** II-1. The innovator chooses randomly an idea among the parents of a randomly chosen idea: $(p_{h,t,1} = 0,\ q_{h,t,1} = 1,\ p_{h,t,2} = 1)$.
3. **Network search**, **Strategy** II-l. The innovator chooses the sequence of actions $(p_{h,t,1} = 0,\ q_{h,t,1} = 1,\ p_{h,t,2} = 0,\ q_{h,t,2} = 1, \ldots, p_{h,t,l+1} = 1)$.

3.2. Network of linkages across ideas generated by given strategies

We now characterize the network $\mathcal{N}_t$ of linkages across ideas formed when innovators use strategies of at most length $l = 1$. Strategies with $l > 1$ are not considered in Lemma 1 (see remark below), but will be considered again in Lemma 2. In order to have a well-defined dynamic process generating the network, we assume that at time 0 there is a set of at least $m + 1$ ideas each of which has at least m parents, with $m > 1$.

First, we need the following definition.

Definition 4. Let $\mathcal{F}(\xi, \beta)$ be the translated power law distribution defined by

$$\mathcal{F}(\xi, \beta) = P[x \leqslant k] = 1 - \left(\frac{\beta}{k + \beta}\right)^{\xi}$$

where $\xi > 1$, $\beta > 0$ and $P[x \leqslant k]$ is the probability that x does not exceed k. Note that the mean of $\mathcal{F}$ is finite for $\xi > 1$ while its second moment is finite only for $\xi > 2$.

Let $\bar{g} = \frac{g}{1+g}$ and denote by k be the number of times an idea has been used in prior knowledge creation, that is, its in-degree.

Lemma 1. *Let* λ, $0 < \lambda < 1$, *be the proportion of innovators choosing Strategy II-1 and let* $(1 - \lambda)$ *be the proportion of innovators choosing Strategy I. Then*:

a) *The network of links across ideas $\mathcal{N}_t$ tends as $t \to \infty$ to a configuration in which the distribution of in-degrees $\Pi(\lambda)$ is stationary and is given for any $k \geqslant 0$ by*[9]

$$p_k = \frac{1}{1+\lambda a}\frac{\Gamma(k+a-1)}{\Gamma(a-1)}\frac{\Gamma(a+\frac{1}{\lambda})}{\Gamma(k+a+\frac{1}{\lambda})}$$

where p_k is the proportion of ideas of in-degree k, Γ denotes the gamma function and $a = (1-\lambda)m$.

b) *According to the "continuum" approximation*[10] *the stationary distribution is a translated power law $\digamma(\xi, \beta)$. The values of the parameters are $\xi = \frac{1}{\lambda}$ and $\beta = \frac{(1-\lambda)m}{\lambda}$ if $\bar{g} = g = 0$, and $\xi = \frac{\log(1+\bar{g})}{\lambda \bar{g}}$ and $\beta = \frac{(1-\lambda)m}{\lambda}$ if $\bar{g} > 0$.*

c) *For pure random search ($\lambda = 0$) the stationary distribution $\Pi(0)$ of in-degrees of ideas is for any $k \geqslant 0$*

$$p_k = \left(\frac{m}{m+1}\right)^k \frac{1}{1+m}$$

Within the "continuum" approximation the distribution is the exponential

$$P\left[k_j^{in} \leqslant k\right] = 1 - \exp\left(-\frac{k}{m}\right)$$

when $g = 0$ while it is $1 - (1+\bar{g})^{-\frac{k}{\bar{g}m}}$ when $\bar{g} > 0$.

d) *For pure network search ($\lambda = 1$) new ideas would never be cited and the dynamic process of knowledge accumulation would break down, unless new ideas have an initial "attractiveness".*[11]

Proof. See Appendix A. □

Lemma 1 indicates that for any environment in which a given proportion $1 > \lambda > 0$ of innovators follow a purely network search, and a proportion $1 - \lambda$ of innovators follow a purely random search, the distribution of in-degrees $\Pi(\lambda)$ is stationary and approximately a power law. Intuitively, ideas with higher in-degree have higher chances of being found through network search, generating a mechanism of "proportional growth" and consequently a power law distribution. The result is similar to Jackson and Rogers [18], although our proof uses the "master equation" approach that provides an exact solution. The "contiunous" approach, chosen by Jackson and Rogers [18], is used here to obtain the more tractable approximation, also reported in Lemma 1.

Lemma 1 also shows that when the proportion of innovators choosing a purely random search strategy tends to zero, implementation of network search becomes difficult as many ideas have zero in-degree, i.e., "too many friends have no friends".

Remark. The process of network formation for long walks ($l > 1$) does not allow an analytical stationary distribution. Evans and Saramaki [12] with numerical analysis obtain the degree distribution for long walks in an undirected network with no growth (see the discussion associated

9 Here we assume $g = 0$ but an analogous expression holds for $g > 0$.

10 See Appendix A for details, or Jackson and Rogers [18].

11 See also Jackson and Rogers [18, Fig. 1 and Remark 25].

with the case $(v\&2) = 1$ and $(v\&4) = 1$ in their Section 4.1 and Fig. 3). They find that the degree distribution is asymptotically a power law of exponent $\xi = 2$.[12]

3.3. Rational behavior for given network

We now characterize the optimal behavior of the innovator when the existing network of linkages across ideas $\mathcal{N}_t$ has a specific in-degree distribution and the distribution of productivity of ideas has a power law upper tail.

Let $\hat{a}_{it}(\tau)$ be the productivity of an idea i created at time t to produce good s and with strategy of length τ. For a given realization of a sub-equilibrium (X_t, E_t), the innovator maximizes the present value of the expected return from innovation net of the search costs. Formally, for given $(\widehat{X}_{it}, E_t)$, where $\widehat{X}_{it}$ is as X_t excluding the productivity of an idea i, the agent h innovating in sector s faces the optimization:

$$Max_{\tau\in\mathbb{N}} E^t\left[V_s\left(\hat{a}_{it}(\tau), \widehat{X}_{it}, E_t\right) - c_{ht}\right]$$

where $E^t[.]$ is the expectation at time t. Let $\lambda_0 = \log(1+\bar{g})/2\bar{g}$. The following existence result is shown to hold.

Lemma 2. *Let λ, $0 < \lambda < \lambda_0$. Let $\mathcal{N}_t$ be such that the stationary distribution of ideas' in-degrees follows a distribution $\Pi(\lambda)$ defined in Lemma* 1. *Assume also that productivities follow a distribution Ξ as in Assumption* 2. *Then there exist strictly positive scalars $\underline{c}_{ht}$, $\bar{c}_{ht}$, $\underline{b_P}$ and $\underline{b_O}$, such that for $b_P < \underline{b_P}$ and $b_O < \underline{b_O}$ the optimal search behavior at time t for an innovator h with search cost c_{ht} is*:

a) *If $\bar{c}_{ht} \leqslant c_{ht}$ the innovator obtains the m parents with strategies of type I.*
b) *If $\underline{c}_{ht} \leqslant c_{ht} < \bar{c}_{ht}$ the innovator obtains the m parents with strategies of type II-1.*
c) *If $c_{ht} < \underline{c}_{ht}$ the innovator obtains the m parents with strategies of type II-l with $l > 1$.*

Proof. See Appendix A. □

Lemma 2 shows that the type and length τ of the optimal search for any given innovator h depends crucially on his marginal (with respect to τ) costs of search. Lemma 2 is a consequence of the assumption that $V_s(\hat{a}_{it}(\tau), \widehat{X}_{it}, E_t)$ is increasing in $\hat{a}_{it}(\tau)$ along any sub-equilibrium. Indeed, for any given state of the economy and time t, the marginal cost of search by assumption is fixed. Consequently, among searches with the same cost the innovator maximizes $\hat{a}_{it}(\tau)$. As productivity $\hat{a}_{it}(\tau)$ depends on the properties of the input parental ideas, the innovator searches for input ideas j having high productivity when considered as inputs, that is, high a_{jt}. The proof reveals that the longer is the search the higher is the expected value of a_{jt}. Note that if $\lambda \geqslant \lambda_0$ then innovators choose strategies of type II-l.

Remark. At time t the appropriate bounds $\underline{c}_{ht}$ and $\bar{c}_{ht}$ depend on the realization of the sub-equilibrium and the identity of the sector assigned to h. If we assume that the strategy decision is independent of the assigned sector (e.g., if that decision is made before the assignment) the innovator would evaluate the expected net returns across all sectors. Assuming the law of large

[12] Evans and Saramaki [12] obtain similar distributions for stochastic long walks with a constant probability of stopping.

number applies, the aggregate variables are approximately deterministic. In this case, the bounds become simple functions of the state of the aggregate variables at time t.

4. Equilibrium path and productivity

The aim of this section is to show that there exists a stationary distribution of the productivity of ideas that is stable for the process of knowledge creation induced by rational innovators. Lemma 1 suggests that, if it exists, such distribution will have a power law upper tail. We focus on stable distributions in the sense that new ideas are created with a productivity that follows the same distribution as that from which the input ideas are extracted. The existence of such a distribution is possible because the productivity of new ideas depends on the productivity of existing ideas and because the network of linkages across ideas is "scale invariant". The latter property requires stationarity of the search strategies along the equilibrium path, which is itself linked to the path of search costs and expected returns from innovation.

We now state the main result of the paper, provide intuition for its proof and, finally, discuss its implications. Let $\bar{g} = \frac{g}{1+g}$ and let k be the number of times an idea has been used in prior knowledge creation.

Theorem 1. *Let $2 < \xi < \infty$. Then there exist $\underline{b_P} > 0$, $\underline{b_O} > 0$ such that for $0 < b_P < \underline{b_P}$, $0 < b_O < \underline{b_O}$ there exist open sets (in the sup-norm) of sequences of individual costs $\overline{\{c_{ht}\}_{t=0}^{\infty}}$ such that there is an equilibrium path $(X_t, E_t)_{t\in\mathbb{N}}$ and*

1. *the upper tail of the distribution of the productivity of the stock of ideas approaches, as $t \to \infty$, a stationary power law of parameter ξ. More precisely, the productivity of ideas follows a distribution $F(z) = P[a_i \leqslant z]$ with unbounded upper support and such that for all $r > 0$*
$$\lim_{z\to\infty} \frac{P[a_i > z]}{P[a_i > rz]} = \lim_{z\to\infty} \frac{1 - F(z)}{1 - F(rz)} = r^{\xi};$$
2. *the endogenous network $\mathcal{N}_t$ of linkages across ideas tends as $t \to \infty$ to a configuration such that the distribution of in-degrees of ideas follows a stationary distribution $\Pi(\lambda)$ as defined in Lemma 1;*
3. *in the "continuum" approximation the stationary distribution of in-degrees is a translated power law $\digamma(\xi, \beta)$ with $\beta = m(\xi - 1)$ if $\bar{g} = g = 0$, and $\beta = m(\frac{\xi\bar{g}}{\log(1+\bar{g})} - 1)$ if $\bar{g} > 0$;*
4. *the value of ξ is related to the proportion λ of innovators searching through networks versus those searching randomly. In particular, when $g = 0$, $\lambda = \frac{1}{\xi}$.*

Proof. See Appendix A. □

An intuition for the proof is as follows. Since $\widetilde{A}_{it}$ and θ_{ijt} are Gaussian variables, Assumption 1 shows that the shape of the distribution relies on the distribution of the number of links attached to an idea, its in-degree. In fact, Assumption 1 implies that if the network of linkages $\mathcal{N}_t$ across ideas is such that the distribution of in-degrees of ideas is a power law with parameter ξ for large degrees, then the distribution of productivities has a power law upper tail, also with parameter ξ. On the other hand, Assumption 2 ensures that when the productivities of ideas are extracted from a distribution with a power law upper tail given by Ξ then there exists a sub-equilibrium $(X_t, E_t)_{t\in\mathbb{N}}$. The second part of the proof requires to show that $(X_t, E_t)_{t\in\mathbb{N}}$ is a long-run equilibrium generated by optimizing innovators. In fact, Lemma 1 ensures that when there is a fixed proportion λ of agents following Strategy II-1 the network of linkages is such that the in-degree follows a displaced power law of parameter ξ. Lemma 2 ensures that, provided the distribution of in-degrees of the network of linkages is a power law with parameter ξ, there exist costs (depending on $(X_t, E_t)_{t\in\mathbb{N}}$) such that a fixed proportion λ of innovators adopt a network search strategy and the remaining a pure random search. Note that if $1 < \xi \leqslant 2$, it is always optimal for innovators to follow network strategies of type II-l (see Appendix A.2.2.2). As discussed in Lemma 1d) a scale-free network is endogenously generated provided new ideas posses initial "attractiveness".

Theorem 1 focuses only on the asymptotic of the distribution of productivities. This is clearly a limitation of the result, but it can be argued that it is not dramatic. Indeed, the aggregate behavior of the economy relies on state-of-the-art techniques and the associated patents. The dynamics of the state-of-the-art is related to the associated extreme value distribution for the largest order statistics.[13] The set of distributions resulting in the Frechet extreme value distribution, called its basin of attraction, is large. It is known that a necessary and sufficient condition for a distribution to be in the basin is that the distribution has unbounded upper support and is asymptotically a power law. The basin is large but does not include popular distributions such as the log normal distribution.

The results of Theorem 1 (i) that the tail of the stationary distribution for the productivity is a power law, (ii) its implication for the largest order statistics, and (iii) that the number of citations follows a displaced power law, all agree with the existing empirical evidence on the productivity of innovations and on patent citations. Theorem 1 also shows that the thickness of the upper tail increases with the relative size of the population of low cost agents but that the existence of high cost agents is necessary to obtain the stable distribution.

Remark. Lemma 2c) shows that for very low search costs, innovators optimally choose longer strategies. We cannot conclude that these would imply a stable stationary distribution because we are not able to obtain an analytical result concerning the modified endogenous network of links. However, as noted after Lemma 1, Evans and Saramaki [12] using numerical methods found that the network converges to a "scale-free" configuration. This suggests that a stable productivity distribution would also exist for very low costs.

Theorem 1 provides the existence result for the equilibrium path and links it to the cost of search. As discussed after Lemma 2, the result highlights the role played by the present value of the returns from innovation and the individual costs of search. Let $E^t[.]$ designate expectations at time t. We then obtain the following result.

Corollary 1. *Along the equilibrium path $(X_t, E_t)_{t=0}^{\infty}$ associated with the stationary distribution of Theorem* 1, *the individual search costs c_{ht} and the expected returns $V_s(\hat{a}_{it}, \widehat{X}_{it}, E_t)$ in sector s are such that $\limsup_{t\to\infty} \frac{E^t[c_{ht}]}{E^t[V_s(\hat{a}_{it}, \widehat{X}_{it}, E_t)]} = \overline{K}_{hs}$ and $\liminf_{t\to\infty} \frac{E^t[c_{ht}]}{E^t[V_s(\hat{a}_{it}, \widehat{X}_{it}, E_t)]} = \underline{K}_{hs}$ for some $0 < \underline{K}_{hs} < \overline{K}_{hs} < \infty$.*

Corollary 1 suggests that the conditions leading to the existence of the invariant distribution of productivity are not as general as one could have expected. This fact is highlighted by the following remark.

[13] In Kortum [22] this is sufficient to obtain the observed long-run trends of productivity and output.

A simple case in which the conditions in Theorem 1 do not hold is given by fixed search costs. Along the equilibrium path and for any $\bar{a} > 0$ the probability density of discovering an idea with productivity $\bar{a}$ is independent of t. Furthermore, we know that for any given productivity $\bar{a}$ of an idea, $\lim_{t\to\infty} V_s(\bar{a}, \widehat{X}_{ht}, E_t) = 0$ because as the state-of-the-art ideas become progressively more productive, eventually the expected revenue associated with $\bar{a}$ is zero. Thus, for constant costs, $c_{ht} = \hat{c}_h$, the optimal strategy of the innovator in the long-run is to abstain from search. We should however stress here that the economic implication of fixed costs has to be taken with care because it relies on the assumption that costs are expressed in working hours (the wage is the numeraire).

The general characterization of the individual cost sequences associated leading to the equilibrium path is difficult because, as noted after Lemma 2, the bounds on the individual costs at time t depend on the state of the economy, including its individual variables. However, suppose it is assumed that innovators evaluate the expected net returns across all sectors and that the law of large numbers ensures that aggregate state variables can be considered as deterministic. Then, if as in Kortum [22], the economy is on a balanced growth path, Corollary 1 implies that on such a path, costs would decrease monotonically.

5. Conclusion

The paper has explored a model in which rational innovators obtain new ideas, resulting in productivity of ideas distributed according to a stationary distribution with a power law upper tail. Innovators are assumed to obtain new ideas from combinations of existing ideas. A stationary distribution for the productivity of ideas then requires the distribution of productivity to be stable for the process of innovation itself. However, this sole condition does not ensure a power law upper tail. The missing “proportional growth” component is provided by the assumption that parent ideas benefit from their use in prior knowledge creation. We have shown that this property may induce the innovator to adopt a behavior that mimics preferential attachment, thus providing the required feedback to “stabilise” the distribution.

The paper has provided insights on the conditions associated with the existence of a stationary productivity distribution with power law upper tail and for the property that the number of times an idea is used follows a displaced power law. The results highlight the importance of the spillover associated with “learning-by-use” and of the locality of the available information, as well as the role of the relative size of the population of innovators with low search costs. Particularly intriguing is the need for both low and high costs agents.

Another insight of this paper is that providing more information on the linkages across ideas may change the incentives of the innovators and affect the endogenous network of linkages across ideas itself. We speculate that this may reduce the spillover that benefits the productivity of input ideas (used for knowledge production) while providing a short term gain in the efficiency of final good production. Eventually, the collapse of the stationary productivity distribution may result, but further research is needed to confirm this.

It has been assumed in this paper that ideas are freely and publicly available, when they are used as inputs. This hypothesis, which implies an extreme intertemporal spillover, is common in the literature, e.g., Aghion and Howitt [3]. However, it can be argued that each innovator knows better the ideas he discovers himself. This may induce the innovator to keep trying to innovate using his own prior ideas. We leave for further research the analysis of this modified model.

Appendix A

A.1. Proof of Lemma 1: Endogenous Network formation for given strategies

As strategies are exogenously assigned to innovators the mathematical problem in Lemma 1 is similar to Jackson and Rogers [18]. However, we prove Lemma 1 using the rate equation approach developed by Krapivsky et al. [23] and Krapivsky and Redner [24] and followed by Vazquez [31], instead of the "continuum approach" used by Jackson and Rogers [18] because it delivers an exact solution. We then use the continuum approach to obtain a simpler and more tractable but approximate solution. Note that in the continuum approach the time dependence of the in-degree is computed assuming that the in-degree k_j of node j is a continuous variable, so that the rate of change of k_j is proportional to the probability of acquiring a new link (this is equivalent to asserting that all agents of the same age have the same degree). Although it is known that this approach does not provide an exact solution, the approximation is very good especially for large values of the in-degree.

A.1.1. Network formation, Case I: $\lambda > 0$ choosing Strategy II-1 and $1 - \lambda$ choosing Strategy I

An innovator choosing Strategy I selects at random exactly m existing ideas. For each of these, the probability that a given existing node j acquires a new attachment in this way is the probability $\frac{1}{N_t}$ that node j is drawn. An innovator choosing Strategy II-1 first selects at random m existing ideas, and then for each of these, say jj, the search continues by selecting one of the m parents of the idea. The probability that node jj is drawn is $\frac{1}{N_t}$ and the probability that the search continues toward j is $\frac{1}{m}$. As j has k_j^{in} such offspring, the probability that j is selected in this way is $\frac{1}{N_t}\frac{1}{m}k_j^{in}$.

Let ΔN_t be the number of new nodes added to the network between time t and $t+\Delta t$. As λR_t chose Strategy II-1 while $(1-\lambda)R_t$ chose Strategy I, the expected number of new links Π_{jt} attached to node j (assuming that there is at most one new link attached to j during Δt) is

$$\Pi_{jt} = m\Delta N_t\left[\lambda\frac{k_{jt}^{in}}{mN_t} + (1-\lambda)\frac{1}{N_t}\right]$$

Assuming, for now $g = 0$, we have $R_t = R_0$ and the expected number of new ideas per period is R_0 (each innovator has one idea per period). Consequently, $\Delta N_t = R_0\Delta t$ and $N_t = tR_0$. The probability of drawing a new node in $[t, t+\Delta t]$ is

$$\Pi_{jt} = \frac{1}{t}\left[\lambda k_{jt}^{in} + (1-\lambda)m\right]\Delta t$$

A.1.1.1. The exact solution. As the expression in the brackets depends on k_{jt}^{in} we denote it by A_k. Let $n_k(t)$ be the number of nodes at time t with in-degree k. Then

$$\Delta n_k(t) = A_{k-1}n_{k-1}(t)\frac{\Delta t}{t} - A_k n_k(t)\frac{\Delta t}{t} + \delta_k\Delta t R_0$$

giving the rate equation

$$\frac{\partial n_k(t)}{\partial t} = \frac{1}{t}\left[A_{k-1}n_{k-1}(t) - A_k n_k(t)\right] + \delta_k R_0$$

where $\delta_k = 1$ if and only if $k = 0$. As we are looking for a stationary in-degree distribution, let $n_k(t) = p_k N_t = t p_k R_0$ where p_k is now the probability that a node has in-degree k. Substitution gives

$$p_k = m\left[\lambda\frac{k-1}{m} + (1-\lambda)\right]p_{k-1} - m\left[\lambda\frac{k}{m} + (1-\lambda)\right]p_k + \delta_k$$

where by assumption $p_{-1} = 0$. For $k = 0$ this equation implies $p_0 = \frac{1}{1+m\lambda(1-\lambda)}$. On the other hand, for all $k \geqslant 1$, we have

$$p_k = \frac{k-1+a}{k+a+\frac{1}{\lambda}}p_{k-1} = \frac{k+a-1}{k+a+\frac{1}{\lambda}}\cdots\frac{a}{1+a+\frac{1}{\lambda}}p_0$$

by recursion where $a = (1-\lambda)m$. As $\Gamma(z+1) = z\Gamma(z)$ we obtain

$$p_k = \frac{1}{1+m\lambda(1-\lambda)}\frac{\Gamma(k+a-1)}{\Gamma(a-1)}\frac{\Gamma(a+\frac{1}{\lambda})}{\Gamma(k+a+\frac{1}{\lambda})}$$

with $a = (1-\lambda)m$. It is known that when a discrete variable k is distributed according to $\frac{\Gamma(k)}{\Gamma(k+\eta)}$ its continuous counterpart is the power law $k^{-\eta}$, which is a good approximation for the asymptotic behavior of the discrete variable k. In the present case, the asymptotic behavior for k is given by $p_k \sim Ck^{-\eta}$ as $k \to \infty$ with $\eta = \frac{1}{\lambda} + 1$. Finally,

$$1 - P\left[k_j^{in} \leqslant k\right] \sim k^{-\frac{1}{\lambda}}$$

A.1.1.2. The approximate solution. Using the continuum approach, as in Jackson and Rogers [18, Lemma 1], from Π_{jt} above we obtain

$$\frac{dk_{jt}^{in}}{dt} = \frac{1}{t}\left[\lambda k_{jt}^{in} + (1-\lambda)m\right]$$

leading to

$$k_{jt}^{in} = m\frac{1-\lambda}{\lambda}\left(\frac{t}{t_j}\right)^{\lambda} - m\frac{1-\lambda}{\lambda}$$

Defining $t(k)$ as the expected date of birth of a node of degree k, we obtain

$$\frac{t(k)}{t} = \left(\frac{\lambda}{(1-\lambda)m}k + 1\right)^{-\frac{1}{\lambda}}$$

The number of agents born before $t(k)$, that is, the population at time $t(k)$, is $R_0 t(k)$. Therefore,

$$1 - P\left[k_j^{in} \leqslant k\right] = \frac{N_{t(k)}}{N_t} = \frac{R_0 t(k)}{R_0 t} = \frac{t(k)}{t} = \left(\frac{\lambda}{(1-\lambda)m}k + 1\right)^{-\frac{1}{\lambda}}$$

A similar analysis holds when the number of innovators grows at a constant rate. In this case, R_t ideas arrive at time t (because innovators produce ideas at a unit rate) so that $\frac{\Delta N_t}{N_t} \simeq (1+g)^t R_0(\frac{(1+g)^{t+1}}{g}R_0)^{-1} = \frac{g}{1+g} = \bar{g}$ where for $g \to 0$, $\bar{g} \to g$. Then

$$\frac{dk_j^{in}(t)}{dt} = \lambda\bar{g}k_{jt}^{in} + (1-\lambda)\bar{g}m = c_1 k_{jt}^{in} + c_2$$

Define d_0 as the in-degree in period t_j, i.e., $k^{in}_{jt_j} = d_0$. Standard manipulations (see, e.g., Jackson and Rogers [17, Lemma 3]) show that the solution to the differential equation obtained above is of the form

$$k^{in}_j(t) = \left(\frac{c_2}{c_1}\right)e^{c_1(t-t_j)} - \frac{c_2}{c_1} + d_0$$

where c_1 and c_2 are the time independent constants defined above. Finally,

$$1 - P\left[k^{in}_j \leqslant k\right] = \left(\frac{c_1}{c_2}k + 1\right)^{-\frac{1}{\lambda \bar{g}}\ln(1+\bar{g})}$$

where $\bar{g} \simeq g$ for small g.

A.1.2. Network formation, Case II: All agents choose Strategy I ($\lambda = 0$)

A.1.2.1. The exact solution. Proceeding as in Case I we obtain

$$\frac{\partial n_k(t)}{\partial t} = \frac{1}{t}m\left[n_{k-1}(t) - n_k(t)\right] + \delta_k R_0$$

where $\delta_k = 1$ if and only if $k = 0$. Using $n_k(t) = p_k N_t = t p_k R_0$ we have

$$p_k R_0 = m p_{k-1} R_0 - m p_k R_0 + \delta_k R_0$$

leading to

$$p_k = \left(\frac{m}{m+1}\right)^k p_0 = \left(\frac{m}{m+1}\right)^k \frac{1}{1+m}$$

A.1.2.2. The approximate solution. Again standard manipulations (Jackson and Rogers [18, Lemma 1]) show that the solution is

$$k^{in}_{jt} = d_0 + m\log\left(\frac{t}{t_j}\right)$$

leading to

$$P\left[k^{in}_j \leqslant k\right] = 1 - \exp\left(-\frac{k}{m}\right)$$

Finally, we consider the case with $g > 0$. The differential equation

$$\frac{dk^{in}_j(t)}{dt} = \bar{g}m$$

admits the solution $k^{in}_j(t) = \bar{g}m(t - t_j) + d_0$. When $d_0 = 0$ we obtain

$$P\left[k^{in}_j \leqslant k\right] = 1 - (1+\bar{g})^{\frac{d_0-k}{\bar{g}m}} = 1 - (1+\bar{g})^{-\frac{k}{\bar{g}m}}$$

Note that when $\bar{g} \to 0$, $(1+\bar{g})^{-\frac{k}{\bar{g}m}} \to \exp(-\frac{k}{m})$, as expected.

A.1.3. Network formation, Case III: All agents choose Strategy II-1 ($\lambda = 1$)

When all agents choose Strategy II-1 and there is no growth, $g = 0$, the relevant differential equation is

$$\frac{dk_{jt}^{in}}{dt} = \frac{1}{t}k_{jt}^{in}$$

Assuming that ideas enter with an initial in-degree d_0, standard manipulations show that the solution is

$$k_{jt}^{in} = d_0\left(\frac{t}{t_j}\right)$$

giving

$$1 - P\left[k_j^{in} \leqslant k\right] = d_0\left(\frac{1}{k}\right)$$

If ideas enter with in-degree equal to zero, $d_0 = 0$, they would never acquire any link and the accumulation process would break down. When there is growth, $g > 0$, the relevant equation is

$$\frac{dk_{jt}^{in}}{dt} = \bar{g}k_{jt}^{in}$$

leading to

$$k_j^{in}(t) = d_0 \exp\left[\bar{g}(t - t_j)\right]$$

Again with $d_0 = 0$ the accumulation process breaks down.

A.2. Proof of Lemma 2: The optimal strategy for a given network

The proof requires computation of the expected value of $\hat{a}_{it}(\tau)$ for a given τ. First, we need to compute the expected productivity of an idea as a function of its position in the network, i.e., its in-degree, and second, to compute the probability that a given search strategy ends on a node of a given in-degree.

A.2.1. Expected productivity of ideas as a function of their position in the network

Assume that b_P is sufficiently small so that the expansion

$$A_t = \sum_{s=0}^{\infty}\left(b_P G_t(\theta_t)\right)^s \widetilde{A}_t$$

is convergent and well defined. Consider component i. The first term of a_{it} is $\widetilde{A}_{it}$, which follows, by assumption, a Normal $N(\overline{A}, \sigma)$. The second order term can be written as

$$\left(G_t(\theta_t)\widetilde{A}_t\right)_i = \sum_{j=1}^{N_t} G_{ijt}(\theta_t)\widetilde{A}_{jt} = \sum_{j\in\mathcal{N}_P(i)} \theta_{ijt}\widetilde{A}_{jt} + \frac{b_O}{b_P}\sum_{j\in\mathcal{N}_O(i,t)} \theta_{ijt}\widetilde{A}_{jt}$$

Each $\theta_{ijt}\widetilde{A}_{jt}$ is the product of two independent random variables distributed according to $N(1,\sigma_\theta)$ and $N(\overline{A},\sigma)$. Since θ_{ijt} and $\widetilde{A}_{jt}$ are independent, we obtain

$$E\big[\big(G_t(\theta_t)\widetilde{A}_t\big)_i\big]=\sum_{j\in\mathcal{N}_P(i)}E[\theta_{ij}]E[\widetilde{A}_j]+\frac{b_O}{b_P}\sum_{j\in\mathcal{N}_O(i,t)}E[\theta_{ijt}]E[\widetilde{A}_{jt}]=m\overline{A}+\frac{b_O}{b_P}k_i^{in}\overline{A}$$

The ith coordinate of the third order term $(G_t^2(\theta_t)\widetilde{A}_t)$ is

$$\begin{aligned}\big(G_t^2(\theta_t)\widetilde{A}_t\big)&=\sum_{h=1}^{N_t}G_{iht}(\theta_t)\left[\sum_{j=1}^{N_t}G_{hjt}(\theta_t)\widetilde{A}_{jt}\right]\\&=\sum_{h\in\mathcal{N}_P(i)}\theta_{iht}\left[\sum_{j\in\mathcal{N}_P(h)}\theta_{hjt}\widetilde{A}_{jt}+\frac{b_O}{b_P}\sum_{j\in\mathcal{N}_O(h,t)}\theta_{hjt}\widetilde{A}_{jt}\right]\\&\quad+\frac{b_O}{b_P}\sum_{h\in\mathcal{N}_O(i,t)}\theta_{iht}\left[\sum_{j\in\mathcal{N}_P(h)}\theta_{hjt}\widetilde{A}_{jt}+\frac{b_O}{b_P}\sum_{j\in\mathcal{N}_O(h,t)}\theta_{hjt}\widetilde{A}_{jt}\right]\end{aligned}$$

Independence implies that $E[(G_t^2(\theta_t)\widetilde{A}_t)_i]$ is

$$m^2\overline{A}+\frac{b_O}{b_P}\sum_{h\in\mathcal{N}_P(i)}k_h^{in}\overline{A}+\frac{b_O}{b_P}k_i^{in}m\overline{A}+\left(\frac{b_O}{b_P}\right)^2\sum_{h\in\mathcal{N}_O(i,t)}\sum_{j\in\mathcal{N}_O(h,t)}\overline{A}$$

where k_i^{in} is the number of offspring to node i, its in-degree. If we assume, to simplify the analysis, that there is no correlation between the degree of neighboring nodes and let the expected degree of a parent be $\hat{k}^{in}$, we obtain that

$$E[a_i]=\overline{A}\big(1+b_Pm+b_Ok_i^{in}+b_P^2m^2+b_Pb_Om\hat{k}^{in}+b_Pb_Ok_i^{in}m+b_O^2k_i^{in}\hat{k}^{in}+\mathrm{o}\big(b^2\big)\big)$$

where $\mathrm{o}(b^2)$ designates an error small compared to b^2. This shows that there exist $\underline{b_P}$ and $\underline{b_O}$ such that for $b_P<\underline{b_P}$ and $b_O<\underline{b_O}$ the expected productivity of an idea i in knowledge production is an increasing function of its in-degree k_i^{in}.

A.2.2. Expected productivity when the network is generated by λN_t agents choosing Strategy II-1 and $(1-\lambda)N_t$ agents choosing Strategy I

As expected, the productivity of an idea i is a function of its in-degree k_i^{in}. Consequently, we need to evaluate the expected k_i^{in}, which depends on the network itself and on the type of search strategy adopted by the innovator. We assume that the network is as in Lemma 1 with $0<\lambda<1$ (we ignore the possibility that $\lambda N(t)$ might not be an integer). In principle, the proof should use the exact distribution for the in-degree as given in Lemma 1a). However, as we will see in the proof the use of the "continuos" approximation is legitimate.

A.2.2.1. Expected in-degree k by an agent following Strategy I. **Case $g>0$.** The probability that a given node h of in-degree k_h^{in} is selected by the search when the agent follows Strategy I is $\Pi(k_h)=\frac{1}{N_t}$. In order to obtain the probability that the search ends on a node of in-degree $k_s^{in}\in(k,k+\Delta k)$, the probability $\Pi(k^{in})$ needs to be multiplied by the number of nodes of degree k^{in} in the interval $(k,k+\Delta k)$, noted $N[k\leqslant k^{in}\leqslant k+\Delta k]$. This number can be obtained from the distribution of in-degrees given previously. We obtain

$$N\big[k\leqslant k^{in}\leqslant k+\Delta k\big]=\xi\left(\frac{b}{a}\right)^{\xi}\left(\frac{1}{k+\frac{b}{a}}\right)^{\xi+1}N_t\Delta k$$

with $a = \lambda\bar{g}$ and $b = (1-\lambda)\bar{g}m$. Dividing the above expression by N_t gives the probability that the search ends on a node of in-degree $k_s^{in} \in (k, k+\Delta k)$. Therefore, the resulting associated p.d.f. is

$$f(k) = \xi\left(\frac{b}{a}\right)^{\xi}\left(k+\frac{b}{a}\right)^{-\xi-1}$$

The expected value of the in-degree of an end-node is

$$\begin{aligned}\langle k_I\rangle &= \xi\left(\frac{b}{a}\right)^{\xi}\int k\left(k+\frac{b}{a}\right)^{-(\xi+1)} dk\\ &= \xi\left(\frac{b}{a}\right)^{\xi+1}\left(\xi(\xi-1)\left(\frac{b}{a}\right)^{\xi}\right)^{-1} = m(1-\lambda)\bar{g}\big(\log(1+\bar{g})-\lambda\bar{g}\big)^{-1}\\ &\simeq m \quad \text{for small } \bar{g}\simeq g\end{aligned}$$

where we have used the fact that

$$\int_0^{\infty} x(c+x)^{-(\xi+1)}\, dx = \left[-\frac{c+\xi x}{\xi(\xi-1)(c+x)^{\xi}}\right]_0^{\infty}$$

with $c = a/b$. The expected value of the in-degree of ideas obtained with Strategy I is then $\langle k_I\rangle \simeq m$ as expected.

Case $g = 0$. Similarly, as

$$P\big[k_j^{in} \leqslant k\big] = 1-\left(\frac{a'}{b'}k+1\right)^{-\xi'}$$

with

$$a' = \lambda, \qquad b' = (1-\lambda)m \quad \text{and} \quad \xi' = 1/a' = 1/\lambda$$

we obtain

$$N\big[k \leqslant k^{in} \leqslant k+\Delta k\big] = \xi'\left(\frac{b'}{a'}\right)^{\xi'}\left(k+\frac{b'}{a'}\right)^{-\xi'-1} N_t\Delta k$$

leading to

$$\begin{aligned}\langle k_I\rangle &= \xi'\left(\frac{b'}{a'}\right)^{\xi'}\int k\left(k+\frac{b'}{a'}\right)^{-(\xi'+1)} dk\\ &= \frac{b'}{a'}\big(\xi'-1\big)^{-1} = \frac{(1-\lambda)m}{\lambda}\left(\frac{1}{\lambda}-1\right) = m\end{aligned}$$

A.2.2.2. Probability density to select a node of in-degree k by an agent following Strategy II-1. The probability that a given node h of in-degree k_h^{in} is selected by the innovator who uses Strategy II-1 to search for m ideas is $\Pi(k_{ht}^{in}) = \frac{1}{N_t}k_{ht}^{in}$. Similarly with the case of Strategy I, we obtain

$$N\big[k \leqslant k^{in} \leqslant k+\Delta k\big] = \xi\left(\frac{b}{a}\right)^{\xi}\left(k+\frac{b}{a}\right)^{-\xi-1} N_t\Delta k$$

Therefore, the probability that the search ends on a node of in-degree $k^{in} \in (k, k + \Delta k)$ is obtained dividing the above expression by N_t. The resulting associated p.d.f. is therefore

$$f(k) = \xi k \left(\frac{b}{a}\right)^{\xi} \left(k + \frac{b}{a}\right)^{-\xi-1}$$

The expected value of the in-degree of ideas found in this way is

$$\begin{aligned}\langle k_{II} \rangle &= \xi \left(\frac{b}{a}\right)^{\xi} \int_0^{\infty} k^2 \left(k + \frac{b}{a}\right)^{-\xi-1} dk = 2\xi \left(\frac{b}{a}\right)^{\xi+2} \left(\xi\left(\xi^2 - 3\xi + 2\right)\left(\frac{b}{a}\right)^{\xi}\right)^{-1} \\ &= 2\left(m\frac{1-\lambda}{\lambda}\right)^2 \frac{1}{(\xi - 1)(\xi - 2)}\end{aligned}$$

provided $\xi > 2$, where we have used the fact that

$$\int_0^{\infty} x^2 (c + x)^{-(\xi+1)}\, dx = \left[-\frac{2c^2 + 2c\xi x + \xi(\xi - 1)x^2}{\xi(\xi^2 - 3\xi + 2)(c + x)^{\xi}}\right]_0^{\infty}$$

Note that if $1 < \xi \leqslant 2$ then $\langle k_{II} \rangle = \infty$. **Case $g = 0$.** Similarly, for agents following Strategy II-1 the expected value of the in-degree of ideas is

$$\begin{aligned}\langle k_{II} \rangle &= \xi' \left(\frac{b'}{a'}\right)^{\xi'} \int k^2 \left(k + \frac{b'}{a'}\right)^{-\xi'-1} dk \\ &= 2m^2 \frac{1-\lambda}{1-2\lambda} > 2m^2\end{aligned}$$

A.2.2.3. Expected productivity for agents following Strategies I and II-1. In A.2.1 we found that the expected productivity $E[a_i]$ is an affine function of the in-degree of an idea i. Therefore, as $\langle k_{II} \rangle$ is larger than $\langle k_I \rangle$ the expected productivity of the idea found with Strategy II-1 is higher than with Strategy I.

A.2.2.4. Expected productivity for long walk strategies II-l, with $l > 1$. We first consider a walk of length $l = 2$ ending on node j. First, we need to evaluate the number of offspring at distance 2. Node j has k_j^{in} offspring at distance 1, noted *jj*. Each *jj* has a k_{jj}^{in} offspring, noted *jjj*. Each of the *jjj* is drawn with a probability $\frac{1}{N_t}$. The probability that the walk passes through *jj* is $\frac{1}{m}$. However, *jj* has k_h^{in} such neighbors. So, the probability *jj* is reached after a length 1 walk is $\frac{k_{jj}^{in}}{mN_t}$. The probability that the walk passing through *jj* ends in j is $\frac{1}{m}$. However, j has k_j^{in} such neighbors. Therefore, the overall probability that node j is selected after a walk of length 2 is

$$\frac{k_{jj}^{in}}{mN_t} \frac{k_j^{in}}{m}$$

With a similar logic for a walk of length n ending on node j we obtain

$$\frac{1}{N_t} \frac{1}{m^n} k_j^{in} k_{jj}^{in} \ldots k_{jj\ldots j}^{in} \ (n \text{ times}) = \frac{1}{N_t} \prod_{s=1}^{n} \frac{k_s^{in}}{m}$$

where the product is over the successive offspring. This result shows that provided $\frac{1}{m}E(k_j^{in}) > 1$ holds for all the visited nodes, increasing the length of the walk increases the probability that high degree nodes are selected. This property is likely to hold as assortativity is at work in the present model.[14] This implies that for very low costs, agents would have incentives to deviate from the short-walk strategy. The networks generated by short walks would then not be stable.

A.2.3. Value of the expected returns from innovation for agents following Strategies I and II-1

From A.2.2 we know that the expected productivity of a parent found with Strategy II-1 is larger than when found with Strategy I. From Assumption 1, we know that this translates to the productivity $\hat{a}_{it}$ of the idea to produce a final good and built on these parents. On the other hand, how the expected productivity of an innovation $E^t[\hat{a}_{it}]$ translates into the expected returns of innovation depends on the state of the economy at time t. We assume that for a given state of the economy these returns are increasing in productivity, i.e., $V_s(\hat{a}_{it}\widehat{X}_{is}, E_s)_{s=t}^{\infty}$ is increasing in $\hat{a}_{it}$. This implies that at time t for each realization of the state of the economy there are bounds on the costs such that it is optimal for the innovator to choose Strategy I or to choose Strategy II-1. Agents with costs within these bounds have no incentive to deviate from the chosen strategy. In general, for each sequence of states of the economy (realizations) there exist sequence of bounds $\{\underline{c}_{ht}\}_{t=0,\,h\in\mathbb{N}_{\approx}}^{\infty}$ and $\{\bar{c}_{ht}\}_{t=0,\,h\in\mathbb{N}_{\approx}}^{\infty}$, such that the innovators have no incentives to change strategy.

A.3. Theorem 1: Distribution of the productivity of ideas

For $b_P < \underline{b_P}$ and $b_O < \underline{b_O}$ the series

$$A_t = \sum_{s=0}^{\infty} \bigl(b_P G_t(\theta_t)\bigr)^s \widetilde{A}_t$$

is convergent and well defined. We focus on the ith component of A_t. The productivity at time t of an idea i used to produce a final good and discovered at time t_i is given by $\hat{a}_{it} = a_{it_i}$ where a_{it_i} is given by Assumption 1 with $\mathcal{N}_O(i, t_i) = \emptyset$ (i.e., an idea i has no offspring at the time of its discovery).

The first term in the series giving a_{it_i} follows from the assumption $\widetilde{A}_{it} \sim N(\overline{A}, \sigma)$. Consider now the second term. The vector $G_t(\theta_t)\widetilde{A}_t$ is such that

$$\bigl(G_{t_i}(\theta_{t_i})\widetilde{A}_{t_i}\bigr)_i = \sum_{j=1}^{N_{t_i}} G_{ijt_i}(\theta_{t_i})\widetilde{A}_{jt_i} = \sum_{j\in\mathcal{N}_P(i)} \theta_{ijt_i}\widetilde{A}_{jt_i}$$

As each innovation has at time of its realization m parents, $(G_{t_i}(\theta_{t_i})\widetilde{A}_{t_i})_i$ is the sum of m random variables. Each variable $\theta_{ijt_i}\widetilde{A}_{jt_i}$ is by assumption the product of two independent random variables extracted from the two Normal distributions $N(1, \sigma_\theta)$ and $N(\overline{A}, \sigma)$. The probability density function of each product $\theta_{ijt_i}\widetilde{A}_{jt_i}$ can be expressed as a series of Bessel functions of the second

[14] The existence of assortativity has been formally proved in the similar model of Jackson and Rogers [18]. The intuition is that high degree nodes tend to be older and tend to have older neighbors, which themselves tend to have higher degree.

kind of a purely imaginary argument (see Craig [10] or Springer [30]). The series representing the density function $F(z)$ is given in Craig [10, Eq. (8)]

$$F(z) = \frac{1}{\pi} e^{-\frac{\rho_1^2+\rho_2^2}{2}} \left[\sum_0 K_0 + (\rho_1^2 + \rho_2^2) \frac{|z|}{2!} \sum_2 K_1 + (\rho_1^4 + \rho_2^4) \frac{|z|}{4!} \sum_4 K_2 + \cdots \right]$$

where K_n are Bessel functions and ρ_1 and ρ_2 the reciprocals of the coefficient of variation, $\rho_1 = \frac{1}{\sigma_\theta}$ and $\rho_2 = \frac{\bar{A}}{\sigma}$. On the other hand, from Mechel [27] we know that for large values of $|z|$, i.e. $|z| \gg 1$, we have

$$K_n(z) \approx \left(\frac{\pi}{2}\right)^{\frac{1}{2}} e^{-z} z^{-\frac{1}{2}}$$

Consequently

$$F(z) \approx \frac{1}{\pi} e^{-\frac{\rho_1^2+\rho_2^2}{2}} \left[\sum_0 \left(\frac{\pi}{2}\right)^{\frac{1}{2}} e^{-z} z^{-\frac{1}{2}} + +(\rho_1^2 + \rho_2^2) \frac{|z|}{2!} \sum_2 \left(\frac{\pi}{2}\right)^{\frac{1}{2}} e^{-z} z^{-\frac{1}{2}} + \cdots \right]$$

which is a very "narrow" distribution as $|z| \to \infty$.

We now consider the third term of the series. The ith coordinate of the vector $G_{t_i}(\theta_{t_i})^2 \widetilde{A}_{t_i}$ is

$$\begin{aligned}\left(G_{t_i}^2(\theta_{t_i})\widetilde{A}_{t_i}\right)_i &= \sum_{h=1}^{N(t)} G_{iht_i}(\theta_{t_i}) \left[\sum_{j=1}^{N(t)} G_{pjt_i}(\theta_{t_i}) \widetilde{A}_{jt_i} \right] \\ &= \sum_{h \in \mathcal{N}_P(i)} \theta_{iht_i} \left[\sum_{j \in \mathcal{N}_P(h)} \theta_{hjt_i} \widetilde{A}_{jt_i} + \frac{b_O}{b_P} \sum_{j \in \mathcal{N}_O(h,t_i)} \theta_{hjt_i} \widetilde{A}_{jt_i} \right]\end{aligned}$$

We know that the number of elements in $\mathcal{N}_O(h, t_i)$ follows a Pareto distribution. Therefore, $\sum_{j \in \mathcal{N}_O(h,t_i)} \theta_{hjt_i} \widetilde{A}_{jt_i}$ is a Pareto stopped-sum of random variables, each random variable having the characteristics described above. Using Propositions 4.3 and 4.9 in Fay et al. [13] it is straightforward to show that $\sum_{j \in \mathcal{N}_O(h,t_i)} \theta_{hjt_i} \widetilde{A}_{jt_i}$ has a Pareto upper tail with the characteristics of the in-degree distribution. Consequently, for each h the product $\theta_{ih} \sum_{j \in \mathcal{N}_O(i)} \theta_{hOj} \widetilde{A}_j$ is the product of two random variables, one extracted from a distribution with a Pareto upper tail and the other from the distribution $F(z)$ defined above. Adapting the arguments in Cline and Samorodnitsky [11], it can be shown that this distribution has a Pareto upper tail. Finally,

$$\sum_{h \in \mathcal{N}_P(i)} \theta_{iht_i} \left[\sum_{j \in \mathcal{N}_P(h)} \theta_{hjt_i} \widetilde{A}_{jt_i} \right]$$

is m times the product of a Normal $N(1, \sigma_\theta)$ and m variable extracted from F. This is again the product of well-behaved random variables with thin upper tail. Asymptotically, the dominant term is

$$b_P^2\left(G_{t_i}^2(\theta_{t_i})\widetilde{A}_{t_i}\right) = b_O b_P \sum_{h \in \mathcal{N}_P(i)} \theta_{iht_i} \sum_{j \in \mathcal{N}_O(h,t_i)} \theta_{hjt_i} \widetilde{A}_{jt_i}$$

The precise asymptotic behavior of this term can be obtained from Fay et al. [13] and from Chap. 8, pp. 268–272 in Feller [14]. Indeed, we know that if a random variable k is distributed according to a p.d.f.

$$g(x) = \frac{\alpha}{\beta} \left(\frac{\beta}{x + \beta}\right)^{\alpha+1}$$

then, as $x \to \infty$, we have

$$P\left[\sum_{i=1}^{m} x_i > x\right] \sim m\left(\frac{1}{1+\frac{x}{\beta}}\right)^{\alpha} L(x)$$

with $\lim \frac{L(\lambda x)}{L(x)} \to 1$ as $x \to \infty$ and for fixed λ, i.e. $L(x)$ is slowly varying at infinity.

With these definitions in place we have $\beta = \frac{c_2}{c_1}$ and $\alpha = \xi$. Therefore, we obtain that

$$\lim_{a\to\infty} \frac{1-F(a)}{1-F(ra)} = \lim_{a\to\infty} \frac{K(1+\frac{a}{\frac{c_2}{c_1}})^{-\xi} L(a)}{K(1+\frac{ra}{\frac{c_2}{c_1}})^{-\xi} L(ra)} = \lim_{a\to\infty}\left(\frac{\frac{1}{a}+\frac{c_1}{c_2}}{\frac{1}{a}+\frac{rc_1}{c_2}}\right)^{-\xi} = r^{\xi}$$

References

[1] D. Acemoglu, D. Cao, Innovation by incumbents and entrants, WP MIT, 2010.
[2] D. Acemoglu, A. Ozdaglar, A. Tahbaz-Salehi, Cascades in networks and aggregate volatility, WP MIT, 2010.
[3] P. Aghion, P. Howitt, A model of growth through creative destruction, Econometrica 60 (1992) 323–351.
[4] A. Alchian, Uncertainty, evolution, and economic theory, J. Polit. Economy 58 (3) (1950) 211–221.
[5] P. Auerswald, S. Kauffman, J. Lobo, K. Shell, The production recipe approach: An application to learning by doing, J. Econ. Dynam. Control 24 (3) (2000) 389–450.
[6] A. Barabasi, R. Albert, Emergence of scaling in random networks, Science 286 (1999) 509–512.
[7] J. Benhabib, A. Bisin, S. Zhu, The distribution of wealth and fiscal policy in economies with infinitely lived agents, WP NYU, 2010.
[8] V. Carvalho, Aggregate fluctuations and the network structure of intersectoral trade, WP, 2010.
[9] J. Cordoba, On the distribution of city sizes, J. Urban Econ. 63 (2008) 177–197.
[10] C. Craig, On the frequency of xy, Ann. Math. Statist. 7 (1) (1936) 1–15.
[11] D.B.H. Cline, G. Samorodnitsky, Subexponentiality of the product of independent random variables, Stochastic Process. Appl. 49 (1994) 75–98.
[12] T.S. Evans, J.P. Saramaki, Scale free networks from self-organisation, Phys. Rev. E 72 (2005) 26138, http://arxiv.org/abs/cond-mat/0411390.
[13] G. Fay, B. Gonzalez-Arevalo, T. Mikosch, G. Samorodnitsky, Modeling teletrafic arrivals by a Poisson cluster process, WP, 2005.
[14] W. Feller, An Introduction to Probability Theory and Its Applications, vol. 2, John Wiley & Sons, New York, 1971.
[15] X. Gabaix, Zipf's law for cities: an explanation, Quart. J. Econ. 114 (3) (1999) 739–767.
[16] X. Gabaix, Power laws in economics and finance, in: Annu. Rev. Economics, Annu. Rev. 1 (1) (2009) 255–294, 05.
[17] G. Grossman, E. Helpman, Innovation and Growth in the Global Economy, MIT Press, Cambridge, 1991.
[18] M.O. Jackson, B.W. Rogers, Meeting strangers and friends of friends: how random are social networks?, Amer. Econ. Rev. 97 (3) (2007) 890–915.
[19] C. Jones, The shape of production functions and the direction of technical change, Quart. J. Econ. 120 (2) (2005) 517–549.
[20] B. Jovanovic, Selection and evolution of industry, Econometrica 50 (1982) 649–670.
[21] M. Kelly, Linkages, threshold, and development, J. Econ. Growth 6 (2001) 39–53.
[22] S. Kortum, Research, patenting, and technological change, Econometrica 65 (6) (1997) 1389–1419.
[23] P.L. Krapivsky, S. Redner, F. Leyvraz, Connectivity of growing random networks, Phys. Rev. Lett. 85 (2000) 4629.
[24] P.L. Krapivsky, S. Redner, Organization of growing random networks, Phys. Rev. E 63 (2001) 066123.
[25] E.G.J. Luttmer, Selection, growth, and the size distribution of firms, Quart. J. Econ. 122 (3) (2007) 1103–1144.
[26] E.G.J. Luttmer, Technology diffusion and growth, J. Econ. Theory 147 (2) (2012) 599–619, this issue.
[27] Fr. Mechel, Calculation of the modified Bessel functions of the second kind with complex argument, Math. Comp. 20 (95) (1966) 407–412.
[28] R.R. Nelson, S.G. Winter, An Evolutionary Theory of Economic Change, Harvard University Press, Cambridge, 1982.

[29] H.A. Simon, On a class of skew distribution functions, Biometrika 42 (1955) 425–440, doi:10.1093/biomet/42.3-4.425.
[30] M.D. Springer, The Algebra of Random Variables, Wiley, New York, 1979.
[31] A. Vazquez, Growing networks with local rules: Preferential attachment, clustering hierarchy, and degree distributions, Phys. Rev. E 67 (2003) 056104.
[32] M. Weitzman, Recombinant growth, Quart. J. Econ. 113 (2) (1998) 331–360.
[33] G.U. Yule, A mathematical theory of evolution based on the conclusions of Dr. J.C. Willis, F.R.S., Philos. Trans. R. Soc. Lond. Ser. B, Containing Papers Biological Character 213 (1925) 21–87.