Journal of Economic Dynamics & Control 52 (2015) 150–165

Contents lists available at ScienceDirect

Journal of Economic Dynamics & Control

journal homepage: www.elsevier.com/locate/jedc

Self-organization of knowledge economies ☆

François Lafond [a,b,c,*]

[a] Institute for New Economic Thinking at the Oxford Martin School, Eagle House, Walton Well Road, Oxford OX2 6ED, UK
[b] London Institute for Mathematical Sciences, 35a South Street, Mayfair, W1K 2XF London, UK
[c] United Nations University - MERIT, Keizer Karelplein 19, 6211TC Maastricht, The Netherlands

ARTICLE INFO

Article history:
Received 11 September 2013
Received in revised form
11 November 2014
Accepted 14 December 2014
Available online 29 December 2014

JEL classification:
D83
D85
O31
O33

Keywords:
Innovation
Diffusion
Two-mode networks
Power law

ABSTRACT

Suppose that homogeneous agents fully consume their time to invent new ideas and learn ideas from their friends. If the social network is complete and agents pick friends and ideas of friends uniformly at random, the distribution of ideas' popularity is an extension of the Yule–Simon distribution. It has a power-law tail, with an upward or a downward curvature. For infinite population it converges to the Yule–Simon distribution. The power law is steeper when innovation is high. Diffusion follows logistic curves.

1. Introduction

The importance of knowledge in explaining economic outcomes has been widely documented. At the individual level, educational training and skills determine income (Schultz, 1961) and capabilities (Sen, 1999). At the firm level, innovation is the source of competitive advantage and profits (Schumpeter, 1934). At the country level, technological change explains most of GDP growth (Solow, 1957).

To understand the process of economic development, one should therefore study the generation and diffusion of ideas. The literature on endogenous growth has significantly clarified the mechanisms through which knowledge can lead to GDP growth (Lucas, 1988; Romer, 1990), but less efforts have been devoted to the study of the detailed distribution of ideas in simple, decentralized "knowledge economies" in which agents create and exchange ideas. Some patterns are more likely or efficient than others (Cowan and Jonard, 2004). For economists, it is crucial to have expectations about the structure of who knows what (the distribution of ideas). For instance, since production relies on knowledge, the structure of who knows what influences the structure of who produces what (product differentiation and countries' specialization). Moreover, since

☆ This paper supplants the relevant parts of "Learning and the structure of citation networks", UNU-MERIT working paper 2012-071, where the model is extended to explain the structure of citation networks.

* Correspondence address: Institute for New Economic Thinking at the Oxford Martin School, Eagle House, Walton Well Road, Oxford OX2 6ED, UK.
E-mail address: francois.lafond@oxfordmartin.ox.ac.uk.

http://dx.doi.org/10.1016/j.jedc.2014.12.004
0165-1889/

knowledge is (mostly) a public good, and is (mostly) cumulative in nature, the structure of who knows what determines and is determined by the rate and direction of inventive activity. Therefore, long-run economic progress depends intimately on the detailed organization of knowledge systems.

This paper analyzes the structure of who knows what by deriving the distribution of ideas' popularity in a simple model based on the assumption that attention is allocated between innovation and imitation. The objectives are first, to find conditions under which a stable system can be characterized, and second, to characterize the resulting organization. In other words, this paper studies the self-organization of knowledge economies: when a collection of agents produces and consumes knowledge, can we expect a certain form of stability in the distribution of knowledge? and if so, which is this stable form? I find that if the trade-off between innovation and imitation is constant, then a stable distribution of ideas' popularity emerges, in spite of the disturbing force of innovation being at play, with new ideas arriving regularly. Moreover, even though which ideas diffuse and which agents are chosen to receive and diffuse knowledge are stochastic events, self-organization produces a certain stability in the average overlap among agents' ideas' portfolios, and hence in the distribution of ideas' popularity. Self-organization can be understood at the mean-field level, where there exists a fixed point, self-consistency equation from which one can derive a steady-state that is unique. In other words, when agents create new ideas and learn random ideas of random friends, after some time the structure of who knows what will be such that the diffusion process is compatible with that same structure, even though it is growing due to innovation. Hence it is a stable, self-organized knowledge economy.

The distribution of ideas' popularity is, roughly speaking, a power law, due to the fact that learning random ideas of random friends produces cumulative advantage (or self-reinforcing dynamics) in ideas' diffusion: the more an idea is known, the higher the chances that it is found at random in a random friend. However, since population is bounded, which ensures bounded diffusion, the power law has finite support (ideas are known by at most the number of agents in the population). When the social network is complete, this finite support power law is characterized precisely, as a discrete distribution which is a particular case of a Generalized Hypergeometric Distribution, and an extension of the Yule–Simon distribution. Changing the social network can change the distribution of ideas' popularity to some extent, and this is investigated mostly using simulations.

The relationship between the special Pfaff–Saalschützian Generalized Hypergeometric distribution derived here and the Yule–Simon distribution follows from the fact that the proposed model can be seen as an extension of Simon (1955) model. In Simon's original model, there are agents and ideas. At each period, a new idea arrives. With some fixed probability b, it goes to a new agent (created simultaneously). Otherwise, it goes to an agent chosen with probability proportional to the number of ideas that he holds. This process leads to a steady-state distribution of the number of ideas per agent which has power law tail, and is called the Yule–Simon distribution. For instance, among many other fields of application, Simon fitted his distribution using scientific authors and their papers. My starting point is that diffusion is missing. Scientific papers, like technologies and social norms, diffuse through the population. For clarity let me abstract from agent heterogeneity, and consider a fixed number of agents. I still want to have a growing number of ideas, consistent with reality, but also wish to allow agents to learn ideas of/from others. Since I contend that attention is limited, I assume that at each period, a randomly

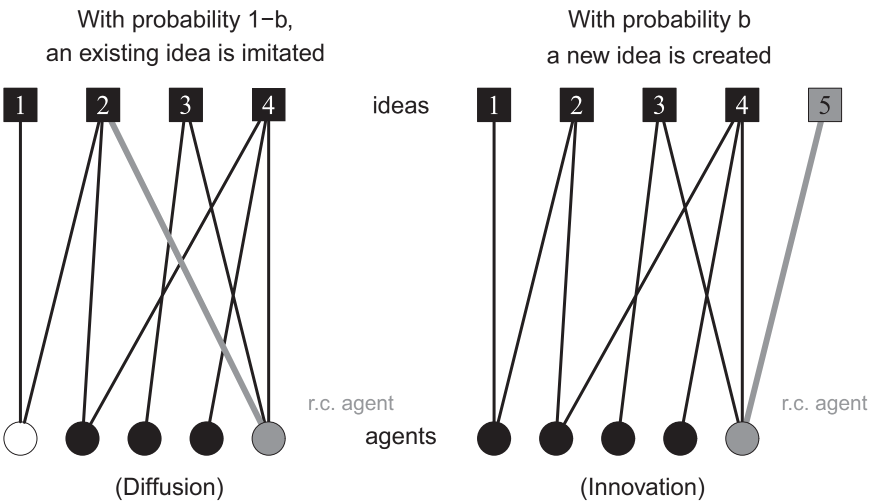

Fig. 1. Schematic description of the model. At each time step, one and only one of the two events represented above happens. In both cases, a link is added. The main focus here is on the degree distribution of the top nodes (ideas' popularity), $p(k)$. The degree distribution of the bottom nodes is discussed in Appendix B but is purposefully uninteresting (agents are homogeneous so it is binomial). On the left panel, the r.c. agent (in gray) is learning. In this case, a neighbor has been randomly chosen and turns out to be the leftmost (in white). There are only two ideas unknown by the r.c. agent and known by the r.c. neighbor (1 and 2). The randomly chosen agent chooses uniformly at random an idea of the r.c. neighbor that he does not know himself— in the example above he turns out to choose the second idea (a link (in gray) is added between the r.c. agent and this idea). On the right panel, the r.c. agent has created a new idea. The social network between bottom nodes, not depicted here, is assumed to be full throughout the paper except in Section 5.1.

chosen agent chooses *either* to innovate, *or* to learn an existing idea. The agent then gets a new edge in the two-mode network of agents and ideas, a (bipartite) network where an edge between agent i and idea j means that "i knows j". The other side of the new edge is either an existing idea or a new one (Fig. 1). As described, the process is close to Simon's, but with one *fixed* set of nodes. This is important because the finiteness of population is necessary for diffusion to be logistic. Simon's master equation for the degree distribution should be modified, using a quadratic instead of linear attachment kernel. The resulting distribution is an extension of the Yule–Simon distribution, and resembles the beta distribution. It converges to the Yule–Simon when the population is infinite. The Yule–Simon distribution has one parameter, which depends on the relative rates of innovation and imitation.

The paper is organized as follows. The next section discusses related literature. Section 3 presents the model and clarifies key mathematical relationships in this setup. Section 4 gives the main results. Section 5 provides some results for two key generalizations (with a sparse social network, and with differentiated productivity of the time spent on imitation or innovation). The last section concludes.

2. Related literature

Cohen and Levinthal (1989) argued that R&D activities allow firms to absorb knowledge spillovers from their environment, reinforcing innovation capabilities. This paper is about the global organization of knowledge systems resulting from the allocation of time between "true" (new-to-the-world) innovation and learning/imitation/diffusion. There is an extensive literature on diffusion (Geroski, 2000), but it generally takes the new technology (idea, product, etc.) as pre-existing, and is concerned only with adoption, elaborating the mechanisms behind diffusion and looking for their idiosyncratic traces in empirical data (Young, 2009). In the information age, much more information is available online, and we could think that individuals and firms learn from a database, instead of from their friends. Why would neighborhood effects in learning be so important then? One reason is that knowledge is tacit, situated, localized or embedded. This stickiness of knowledge implies that it can diffuse only, or preferably, face-to-face (Breschi and Lissoni, 2009). Relatedly, social embeddedness channels awareness of ideas: one may learn new knowledge from a book or online after having been referred to it (by a peer). An important consequence of word-of-mouth interaction is that the diffusion pattern is likely to be logistic, in agreement with the literature on diffusion (Mansfield, 1961). In fact, learning from others naturally introduces increasing returns in ideas' diffusion due to the fact that well-known ideas have more chances to diffuse, because they have more carriers. In the model below, as in the literature, growth is constrained by population size in such a way that the diffusion is logistic. Logistic diffusion is well established theoretically and empirically, which leads to the two following questions: (i) What happens when there are *many* ideas competing for attention? (ii) What happens when there is continuous arrival of *new* ideas?

A way to characterize a system in which ideas are created and diffuse is by keeping track of the distribution of ideas' popularity.[2] In the language of networks, this is the degree distribution of the "ideas" set of a two-mode network of agents and ideas. I assume one fixed set of nodes (the number of agents does not change) and one growing set of nodes (the number of ideas increases without bound). This framework allows to keep track of who knows what in a very detailed way, and provides a bridge between social network models (a one-mode network of agents) and epistemic network models (a one-mode network of ideas). Such a representation of the co-evolutionary dynamics of social and knowledge networks has been used in empirical studies (Roth and Cointet, 2010) and simulation modelling (Börner et al., 2004). Cowan and Jonard (2009) study a closely related system, where firms form an alliance network based on knowledge matching. In their model, firms hold ideas and form pairwise alliances with other firms in order to innovate. Partner choice is based on knowledge overlap: too much overlap would mean that partners have few things to learn from each other; too little overlap may hinder mutual understanding. Their model reproduces several empirical facts of R&D networks, such as small world properties and skewed degree distribution.

The model developed here is also closely related to models of network growth based on copying (Vázquez, 2003). For instance, in Jackson and Rogers's (2007) model for social networks, newborn agents choose to link to random existing agents (random meetings), and to random neighbors of their random meetings (search). Here, cumulative advantage comes from search meetings, because the more friends an agent has, the higher the chances he has to be found through a friend. Likewise, in a two-mode network, search can generate cumulative advantage and, ultimately, fat tail distribution of popularity. This was clearly demonstrated by Evans and Plato (2008), who considered a fixed set of agents and a fixed set of artefacts. Agents are connected to an artefact, and, when they are chosen, connect to another artefact by imitating a friend. Their model is a two-mode network with both sets of nodes fixed, and a rewiring process. Actors are linked to one and only one artefact, and the distribution of artefacts' popularity is studied. Their model applies, for instance, in anthropology where one is interested in the transmission of cultural artefacts. The model proposed below also applies to this context, but assumes that new artefacts appear over time, and that actors accumulate artefacts over time.

[2] For models of knowledge growth and diffusion which do not involve networks, see e.g. Jovanovic and Rob (1989), König et al. (2012) and Lucas and Moll (2014). The model presented here is complementary, because these models are more elaborated in terms of agents' choices and economic observables (e.g. GDP or productivity), but my model is richer in terms of the underlying combinatorial structure. For instance, since ideas are discrete in the model below, two agents with the same number of ideas can imitate ideas of each other, whereas two agents with the same productivity level cannot learn from each other in e.g. Lucas and Moll (2014).

Another closely related model was studied by Ramasco et al. (2004). As Simon, they considered only the production of ideas (papers) but the number of agents is allowed to grow and papers are co-authored. Their work focused on reproducing the empirical data on the "co-authorship" network. Assuming that authors are chosen for new authorship with probability proportional to the number of their previously authored papers, Ramasco et al. (2004) derive the Yule–Simon distribution (with modified parameters) for the distribution of the number of papers authored by an author, and a shifted power law for the degree distribution of the co-authorship network. There have been other studies of two-mode or multi-mode networks in which all sets of nodes are growing. Peruani et al. (2007) studied a model in which only one set of nodes is growing, as in the model presented here, but they analyze the degree distribution of the non-growing side, while the focus here is on the growing side. Beguerisse-Díaz et al. (2010) studied a system in which users rate videos. Liu et al. (2011) study a social tagging system, which can be seen as a three-mode network (users, resources tagged, and tags). Zeng et al. (2012) show that certain recommender systems produce more unequal popularity distribution than others.

The model proposed below contributes to the literature on "self-organizing" networks by providing a detailed analysis of the artefact degree distribution under the assumption of a non-growing population of actors and assuming a specific one-mode network for agents' interactions. Technically, the most noticeable feature of the model is that the probability for a given idea to diffuse at time t (the attachment kernel) is a *quadratic* function of its popularity at time t. This gives rise to a combinatorial interpretation of the partition factor of this attachment kernel. In the classical growing network model with sub-linear attachment kernel (Krapivsky et al., 2000), the value of the partition factor of the attachment kernel cannot be solved for in closed-form, and is computed numerically. In the model below, such a solution may exist but it is hard to find as it involves solving a polynomial of order n (number of agents).

More generally, the model presented here relates to a larger literature on self-reinforcing processes in economics (Arthur, 1989; Kirman, 1993). Following Simon's own applications of his model, notably to the size distribution of firms (Ijiri and Simon, 1977), there has been a large literature impossible to review here. In an influential contribution, de Solla Price (1976) applied Simon's process to explain the power law distribution observed for the in-degree of citation networks. He assumed that existing papers are cited with probability proportional to the number of citations that they have already received. This assumption can be microfounded, by assuming that papers are found by searching through the bibliography of other papers (Vázquez, 2003). The model below allows for an alternative microfoundation of citation networks. In a related paper (Lafond, 2014), using the model described below, and assuming that (an infinite number of) agents cite papers chosen uniformly at random among the papers that they have previously learned or written, the predicted citation distribution is a shifted power law.

3. The model

Before turning to the technical presentation of the fully fledged model, it is useful to present the main assumptions and results as follows:

Assumption 1 (*Knowledge growth and innovation*). Knowledge is a set of discrete ideas. This set is expanding because new ideas are invented over time.

Assumption 2 (*Social embeddedness and diffusion*). Agents imitate ideas of their friends. More precisely, agents choose uniformly at random (u.a.r.) an (unknown) idea of a friend chosen u.a.r.

Assumption 3 (*Limited attention and innovation/imitation trade-off*). Homogeneous agents supply inelastically a fixed amount of attention to obtain ideas. Because some ideas must be invented (Assumption 1), and some must be imitated (Assumption 2), attention is split between these two activities. I assume that this split is the same for all agents and is constant over time.[3]

Result 1. *Social embeddedness creates cumulative advantage for ideas' diffusion, that is, if diffusion was unbounded, ideas would diffuse at a rate proportional to their current popularity.*[4] *However, diffusion is constrained by population size, as in logistic diffusion models.*

Result 2. *This logistic diffusion of sequentially created ideas gives rise to a steady-state distribution of ideas' popularity which is close to a power law but with an upward or a downward curvature in the tail. This curvature disappears when $n \to \infty$ and the distribution is the Yule–Simon distribution. A higher share of attention devoted to innovation (respectively, imitation) generates a steeper (flatter) power law.*

Since the model is stochastic, and because the boundedness of the population introduces a nonlinearity, the derivations of the steady-state distribution are rather involved and can obscure the gist of the argument. Hence, before going into the details of the stochastic model, let me first present a simplified version of the model that allows to see why learning ideas of

[3] I regard the innovation/imitation choice as exogenous, because the forces determining choice can be modelled *independently*, that is, there exist several choice theories compatible with the innovation/diffusion process that I describe.

[4] Throughout the paper, the popularity of an idea is the number of times it is known, that is, the number of agents who have adopted/learned/imitated it.

friends under an innovation/imitation tradeoff gives rise to a power law distribution of ideas' popularity. The derivations are a two-mode version of Barabási and Albert's (1999) original procedure. Assume an infinite population and a deterministic diffusion. Once an idea is invented, it diffuses. Since agents learn ideas of their friends, the more carriers an idea has, the more chances it has to diffuse. So it diffuses at a rate proportional to its popularity. However, it competes with all other ideas, which also diffuse at a rate proportional to their popularity – so to write the diffusion rate it will be necessary to divide the popularity of each idea by the "total popularity" in the system. If exactly one agent–idea relationship is added per time period, the total of all popularity is the number of periods, t. Hence k_j, the popularity of idea j born at time t_j, evolves as follows:

$$\dot{k}_j(t) = (1-b)k_j(t)/t.$$

The factor $(1-b)$ has been added because I assume that a fraction b of time is spent on innovation, which limits the speed of diffusion. Using the initial condition $k_j(t_j)=1$ (j is invented by one agent, at some time t_j) this differential equation has solution

$$k_j(t) = \left(\frac{t}{t_j}\right)^{1-b}. \tag{1}$$

Knowing when ideas are born and their popularity, one can tell, at any point in time, how many of them have a certain popularity. Indeed, the share of ideas known k times, denoted $p(k)$, can be found by starting from the cumulative distribution function:

$$\Pr(k_j \le k) = \Pr\left(\left(\frac{t}{t_j}\right)^{1-b} \le k\right) = 1-\Pr\left(t_j \le tk^{-1/(1-b)}\right).$$

Assume that ideas arrived sequentially, in such a way that the t_j's are uniformly distributed, i.e. $\Pr(t_j = Y) = 1/t$ for Y from 1 to t, so $\Pr(t_j \le Y) = \sum_1^Y 1/t = Y/t$. This leads to $\Pr(k_j \le k) = 1-k^{-1/(1-b)}$. Apply $p(k) = dPr(k_j \le k)/dk$ to retrieve the probability distribution of ideas' popularity:

$$p(k) = \hat{b}k^{-1-\hat{b}},$$

where $\hat{b} = 1/(1-b)$. It is easy to check that $\int_1^\infty p(k)\,dk = 1$. This is a power law which steepens with b. The power law exponent is best rewritten as $\gamma = 2+b/(1-b)$ to show that for $0<b<1$ it is greater than 2, and depends positively on the ratio of the share of innovation over the share of diffusion.

The heuristic description above does not account properly for the finiteness of the population, and therefore fails to feature a logistic diffusion pattern (see Eq. (1)). It does not include the structure of social interactions, and is deterministic. I describe below a more complete mathematical model and its numerical (agent-based) simulation.

3.1. The algorithm

Consider a two-mode network with n agents and w ideas. Ideas are either known or unknown by any given agent, which is represented by the presence or the absence of a link between an agent and an idea. The number of agents is kept fixed, but the number of ideas grows. Time is discrete and indexed by t. Denote by E_t the total number of agent–ideas relationships, i.e. the number of edges of the two-mode network. At the beginning ($t=1$), there is one idea known by one randomly chosen (r.c.) agent ($w_1 = 1$ and $E_1 = 1$). Then at each time period, as illustrated in Fig. 1, the following algorithm is applied (where random always means uniformly at random):

I. Pick an agent i at random.
II. With probability b, the agent i creates a new idea (a new node is added to the set of ideas, and an edge is added to the two-mode network, between i and the new node);
III. otherwise (i.e. if the r.c. agent does not create a new idea), pick another agent i' at random. Then pick at random an idea j among those ideas known by i' and unknown by i. Then i learns j (an edge is added to the two-mode network, between i and j).[5]

The following section clarifies the setup of the model by deriving key mathematical relationships implied by the algorithm (I–III).

[5] If both b and n are very small, there are not enough new ideas to satisfy the number of required learning events. This problematic configuration always happens with non-negative probability, and to ensure that the model always run, the computer code is as follows: when a r.c. agent i is supposed to learn but his chosen neighbor has nothing new, i creates a new idea. Again, there will always exist a positive probability to find a (directed) pair that cannot perform the exchange. This probability is small in the region of interest, so I do not include this effect in the derivations. In particular, I consider that a knowledge economy is *defined* for $\mu > 0$ (μ is an increasing function of b and n to be defined later. See *infra* and Fig. 3). Moreover, one can correct the main theoretical result equation (10) simply by using the "empirical" (from the simulation) values of $b = w_t/E_t$ and μ (Eq. (4)). This condition $\mu > 0$ illustrates that there cannot exist a knowledge economy in which, at a global level, ideas are imitated faster than they are created. This constraint is due to the

3.2. Preliminary results

This section derives two results which will be necessary later and help understanding the setting. First, it is explained that the network density tends to a constant in the long run, which is key to obtain a steady-state distribution, and makes clear the importance of assuming a constant b. Second, an expression is derived for the probability that a r.c. idea is known by two r.c. agents. This will be useful when deriving the chances that an idea is unknown by a r.c. agent but known by his friend, which essentially determines the speed of diffusion of individual ideas and hence the popularity distribution.

Consider a matrix Q which has a fixed number of rows (n agents) and a number of columns that depend on time (w_t ideas). The entries Q_{ij} are equal to one if agent i knows idea j, and zero otherwise. This matrix is the incidence matrix of the two-mode network where agent i is linked to idea j if and only if agent i "knows" idea j. Start at $t=1$ with a column vector filled with a one and $(n-1)$ zeros. At each period, with probability b, a column is added (a new idea is created). Then, with probability 1, one entry of Q is changed from zero to one (if a new column has been added, this modified entry must be in that new column). Since exactly one 1 is added at each period, the total number of ones in Q, which is the total number of edges in the network, is $E_t = t$. The total number of ideas w_t is a random variable equal to W if there has been exactly $W-1$ successes out of $t-1$ trials, success happening with probability b. Hence the expected number of ideas is $E(w_t) = 1-b+bt$. Throughout the paper the concern will be on the long run equilibrium state of the system so I will use $w_t = bt$. Then, it is direct to see

Lemma 1. *The density of the system, defined as the two-mode network density and denoted D, is stable in the long run*:

Proof.
$$D_t = \frac{E_t}{nw_t} \approx \frac{1}{nb} \qquad \square$$

Lemma 1 shows that if fluctuations due to the stochastic nature of w_t are omitted (which is legitimate in the long run), the density of the two-mode network is constant (independent of system time t), since b and n are fixed parameters. Time independence of the two-mode network density suggests that there may exist a steady-state degree distribution. Lemma 1 shows that an increased rate of innovation b will make the system sparser (since there are more ideas and agents are learning less often), whereas a high rate of learning $(1-b)$ will make it denser. In this model, growth corresponds to the increment of a column. Diffusion ensures that the density of the system stays stable, by adding positive entries in existing columns.

The key to characterize the self-organized steady-state of the system is to find the number of ideas shared by two r.c. agents, that is, the number of common ideas in a r.c. pair. This is because diffusion takes place between two agents, and is conditioned by what both agents know, since an agent learns only something that his neighbor knows but that he does not know himself. Denoting by N_i the set of ideas of agent i and by $|N_i|$ its cardinal, we have

Lemma 2. *Consider all pairs of agents* (i, i') *in a system with n agents and w ideas. Then the average* (*over all pairs*) *of the number of ideas known by both i and i′ is*

$$\langle |N_{i'} \cap N_i| \rangle = \frac{\sum_{i'<i} |N_{i'} \cap N_i|}{\#\ of\ pairs} = \frac{\sum_{k=1}^{n} \binom{k}{2} P(k)}{\binom{n}{2}} = \frac{w(\langle k^2 \rangle - \langle k \rangle)}{n(n-1)} \tag{2}$$

Proof. Observe that the sum over all pairs of $|N_{i'} \cap N_i|$ is simply the total number of pairwise "overlaps" in the system, i.e. the total number of times that the triplet "two agents linked to an idea" can be found in the network. Since each idea known k_j times produces $\binom{k_j}{2}$ overlaps between pairs, and denoting $P(k)$ the number of ideas with degree k, the sum can be obtained. Using $P(k) = wp(k)$ and denoting $\langle k^r \rangle = \sum_{j=1}^{w} k_j^r = \sum_{k=1}^{n} k^r p(k)$ gives the simplified form. $\square$

Note that Lemma 2 holds in quite general conditions but gives only the average value of pairwise overlap, not its distribution across different pairs. The average will be very informative because the distribution turns out to be tightly peaked around its mean, since there are no specific sources of agents' heterogeneity. In practice, Lemma 2 will often be used after substituting $\langle k \rangle = E_t / w_t = 1/b$.

The main objective is to derive $p_t(k)$, the probability that a r.c. idea in t is known k times (i.e. has degree k). Under what conditions will idea j be learned at time t? First, the r.c. agent i must be learning, which happens with probability $(1-b)$. Second, the r.c. pair (i, i') must be such that $j \in N_{i'} \backslash N_i$, that is, j belongs to the set of ideas which are both known by i' and not known by i. Third, idea j must be the one chosen among all other ideas $j' : j' \in N_{i'} \backslash N_i$. At each period, conditional on the event "learning" being realized, exactly one idea must be chosen. The attachment kernel gives the probability that a particular one

(*footnote continued*)
assumption of inelastic supply of (cognitive) labor. Still, it is possible for individual agents to imitate faster than they innovate, because one newly created idea can be imitated $(n-1)$ times.

be chosen, that is

$$A_t(k_j) := \Pr(k_j(t+1) = k_j(t)+1); \qquad \sum_{j=1}^{w_t} A_t(k_j) = 1-b$$
$$A_t(k_j) = (1-b)\left\langle \frac{\Pr(j \in N_{i'} \backslash N_i)}{|N_{i'} \backslash N_i|} \right\rangle,$$

where the angle brackets denote average over all possible (ordered) pairs. Since agents and pairs of agents are ex ante homogeneous, and since there is no source of strong ex post heterogeneity, I simply assume that all pairs have the same value of $\Pr(j \in N_{i'} \backslash N_i)$ and $|N_{i'} \backslash N_i|$. Hence,

$$A_t(k_j) = (1-b)\frac{\Pr(j \in N_{i'} \backslash N_i)}{|N_{i'} \backslash N_i|}. \tag{3}$$

The probability that $j \in N_{i'} \backslash N_i$ can be computed as follows. We want to know the number of choices for finding an ordered pair (i, i') such that $j \in N_{i'} \backslash N_i$. There are k_j choices for i' such that $j \in N_{i'}$ and $(n-k_j)$ choices for i such that $j \notin N_i$. Hence there are $k_j(n-k_j)$ choices for an ordered pair such that $j \in N_{i'} \backslash N_i$. Since there are $n(n-1)$ ordered pairs, a r.c. pair will exhibit $j \in N_{i'} \backslash N_i$ with probability $k_j(n-k_j)/n(n-1)$. The denominator $|N_{i'} \backslash N_i|$ can be computed as $|N_{i'} \backslash N_i| = |N_{i'}| - |N_{i'} \cap N_i|$. Using again the assumption that this is the same for all pairs, $|N_{i'} \backslash N_i| = \langle |N_{i'}| \rangle - \langle |N_{i'} \cap N_i| \rangle = \mu t/(n-1)$, where Eq. (2) was used and μ is defined as

$$\mu(t) := 1 - \frac{w_t}{E_t}\frac{\langle k^2 \rangle}{n} = 1 - \frac{\langle (k/n)^2 \rangle}{D_t}. \tag{4}$$

The last step in Eq. (4) uses $w_t = bt, E_t = t$ and Lemma 1. I omit the time subscript in $\langle k^2 \rangle = \sum_{j=1}^{w_t} [k_j(t)]^2$. The attachment kernel equation (3) can now be written as

$$A_t(k_j) = \frac{k_j(n-k_j)}{\hat{b}\mu n t}. \tag{5}$$

The condition $\sum_{j=1}^{w_t} A_t(k_j) = 1-b$ is the same equation as the definition of μ (Eq. (4)). μ ensures that the attachment kernel is correctly normalized, that is, if the event of period t is imitation, the chances that a particular idea diffuses are such that exactly one will diffuse. In this sense, μ characterizes the degree of competition among ideas. The higher the μ, the lower the chances that each particular idea diffuses. μ indicates how many ideas are available for diffusion, in a precise sense. Since the chances of "meeting" an unknown idea j is the number of times that j is known by somebody else (or by a friend, if the friendship network is sparse), at this level each idea competes with all ideas *unknown* by a r.c. agent (not with all other ideas in the system). Algebraically, μ as defined in Eq. (4) admits the following combinatorial interpretation:

Proposition 3.1. *μ is the average of the individual quantities μ_i, where μ_i is the fraction of edges that are pointing to ideas unknown by agent i:*

$$\mu = \frac{1}{n}\sum_{i=1}^{n} \mu_i; \quad \mu_i = \frac{\sum_{j \notin N_i} k_j}{\sum_{j=1}^{W} k_j} \tag{6}$$

Proof. The denominator of μ_i is simply the total number of edges, E_t. The numerator of μ_i can be rewritten as $\sum_{j \notin N_i} k_j = \sum_{j=1}^{W} k_j(1-Q_{ij})$ where Q_{ij} are the entries of the incidence matrix, equal to one if i knows j and zero otherwise. Hence,

$$\mu = \frac{1}{nE_t}\sum_{i=1}^{n}\sum_{j=1}^{w}\left[k_j(1-Q_{ij})\right].$$

Transposing the two sums and decomposing the sum over i, this becomes

$$\mu = \frac{1}{nE_t}\sum_{j=1}^{w}\left[\sum_{i=1}^{n} k_j - \sum_{i=1}^{n} k_j Q_{ij}\right].$$

It is easy to see that by definition $\sum_{i=1}^{n} k_j = nk_j$ and $\sum_{i=1}^{n} k_j Q_{ij} = k_j^2$. Therefore,

$$\mu = \frac{1}{nE_t}\sum_{j=1}^{w}\left[nk_j - k_j^2\right] = 1 - \frac{w_t}{E_t}\frac{\langle k^2 \rangle}{n} \qquad \square$$

The factor $\mu(t)$ is defined at all periods of time and helps characterizing the dynamics of the system. However, it depends itself on the dynamics of the system. How is this feedback loop solved? Does the system stabilize? Since the distribution $p(k)$ depends on the attachment kernel, and the attachment kernel depends on μ which depends on the second-order moment of the distribution, Eq. (4) is a fixed point equation, i.e. $\mu = f(\mu, b, n, t)$. If the popularity distribution is stable, its second order moment is stable and so is μ.

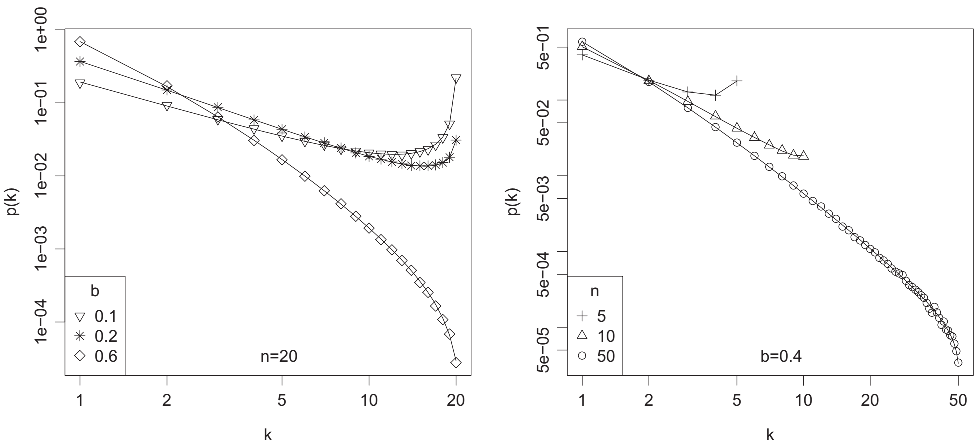

Fig. 2. Distribution of ideas' popularity. For each of six configurations of parameters, the model is run once for 3×10^6 periods (only 10^6 when $n=5, n=10$). In the left panel, $n=20$ and the effect of b is studied. In the right panel, $b=0.4$ and the effect of n is studied. The solid lines are the theoretical results, computed using Eq. (10) and values of μ computed using the fixed point equation (13). These six points of the parameter space are marked in Fig. 3. When a point in Fig. 3 is in the lower left half of the (μ, b) plane, the corresponding curve in the figures above exhibits an upward curvature, otherwise it exhibits a downward curvature.

I show below that assuming that μ is constant and that a steady-state exists, the steady-state is unique. This gives a steady-state value of $\langle k^2 \rangle$, which can be inserted into Eq. (4) to obtain a steady-state fixed point equation for μ.

4. Results

4.1. Distribution of ideas' popularity

In view of the attachment kernel (5), the flows in and out of the kth bin of the histogram can be written explicitly, following the method of Simon. Recall that $P_t(k)$ is the total number of ideas with degree k at time t. Then,

$$P_{t+1}(k) - P_t(k) = P_t(k-1)A_t(k-1) - P_t(k)A_t(k).$$

Using $P_t(k) = btp_t(k)$ and $A_t(k)$ from Eq. (5),

$$t\left(p_{t+1}(k) - p_t(k)\right) + p_{t+1}(k) = p_t(k-1)\frac{(k-1)(n-(k-1))}{\hat{b}\mu n} - p_t(k)\frac{k(n-k)}{\hat{b}\mu n}.$$

Assuming a steady state in the sense that $p_{t+1}(k) = p_t(k) = p(k)$ gives the recurrence

$$p(k)(k(n-k) + \hat{b}\mu n) = p(k-1)(k-1)(n-(k-1)). \tag{7}$$

Eq. (7) can be iterated to give

$$p(k) = p(1)\prod_{i=1}^{k-1} \frac{i(n-i)}{\hat{b}n\mu + (i+1)(n-(i+1))}. \tag{8}$$

Making use of the quadratic formula, the denominator can be rewritten as $(-1)(i-u_1)(i-u_2)$ where $\{u_1, u_2\} = 1/2(n-2 \pm \sqrt{n(n+4\hat{b}\mu)})$. Now consider the definition of the Pochhammer symbol:

$$(x)_y = x(x+1)(x+2)\ldots(x+y-1) = \frac{\Gamma(x+y)}{\Gamma(x)}. \tag{9}$$

Expanding the product in (8) and using (9) on each of the terms give

$$p(k) = p(1)\frac{(1)_{k-1}(n-(k-1))_{k-1}}{(-1)^{k-1}(1-u_1)_{k-1}(1-u_2)_{k-1}}.$$

From Slater's (1966, formula I.5 p. 239), $(n-(k-1))_{k-1} = (-1)^{k-1}(1-n)_{k-1}$. Therefore,

Proposition 4.1. *The steady-state distribution of ideas' popularity is given by*

$$p(k) = p(1)\frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(r_2)_{k-1}}, \tag{10}$$

where

$$\{r_1, r_2\} = \frac{4-n \pm \sqrt{n(n+4\hat{b}\mu)}}{2}$$

and

$$p(1) = \left(1 + \frac{n-1}{n\hat{b}\mu}\right)^{-1}. \tag{11}$$

The term $p(1)$ is found by setting up the appropriate master equation in which there are no inflows from the 0th bin but there is a probability of innovation: $P_{t+1}(1) - P_t(1) = b - P_t(1)A_t(1)$. Assuming a steady-state and solving for $p(1)$ give (11).

The probability mass function (10) is plotted against simulations in Fig. 2. In some region of the parameter space, it has an upward curvature in the tail.[6] This curvature exists when the function admits a minimum at some $k = k^* < n$. Using (10), the point at which $p(k^*) = p(k^* - 1)$ is given by $k^* = \frac{1}{2}\left(1 + n\left(1 + \hat{b}\mu\right)\right)$ and the point at which $p(k^* + 1) = p(k^*)$ is given by $k^* = \frac{1}{2}\left(-1 + n\left(1 + \hat{b}\mu\right)\right)$ so that we may take $k^* = \frac{1}{2}n\left(1 + \hat{b}\mu\right)$. The condition $k^* < n$ is then the same as $\mu < 1 - b$. The region of the parameter space for which this condition holds, such that an upward curvature exists, is the lower left half of Fig. 3 (see Section 4.2). The latter corresponds to relatively low values of b and n (but conditional on b and n being large enough to have $\mu > 0$; see footnote 5).

To obtain further insights onto the nature of the distribution (10), consider verifying that the terms sum up to one. These terms are hypergeometric, so the sum is of the form

$$\sum_{k=1}^{n} p(k) = p(1) \sum_{k=1}^{n} \frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(r_2)_{k-1}} = p(1)\,{}_3F_2[\{1,1,1-n\},\{r_1,r_2\},1].$$

The five parameters of this generalized hypergeometric function (${}_3F_2[\ldots]$) satisfy an important constraint. This ${}_3F_2$ is 1-balanced, that is, its parametric excess is equal to one:

$$(r_1 + r_2) - (1 + 1 + (1-n)) = 1.$$

It means that this ${}_3F_2$ is Saalschützian. Hence, the Pfaff–Saalschütz summation theorem can be applied to check that (10) and (11) define a properly normalized probability mass function

$${}_3F_2[\{1,1,1-n\},\{r_1,r_2\},1] = \frac{(r_1-1)_{n-1}(r_1-1)_{n-1}}{(r_1)_{n-1}(r_1-2)_{n-1}} = \frac{n(1+\hat{b}\mu)-1}{\hat{b}\mu n} = 1/p(1).$$

Note that many other distributions are, in this sense, Pfaff–Saalschützian. More generally, the steady-state distribution (10) is a generalized hypergeometric probability distribution (GHPD). It is named so because its generating function is a ratio of generalized hypergeometric functions (Johnson et al., 2005). In the case of (10), the generating function takes the following particular form[7]:

$$G(z) = \sum_{k=1}^{n} p(k) z^k = \frac{{}_3F_2[\{1,1,1-n\},\{r_1,r_2\},z]}{{}_3F_2[\{1,1,1-n\},\{r_1,r_2\},1]}.$$

This class is interesting because there exists a deep connection between Pfaff–Saalschütz and Gauss hypergeometric theorems. Slater (1966, p. 4849) shows how Gauss theorem can be obtained starting from Pfaff–Saalschütz theorem, and Johnson et al. (2005) show that Gauss hypergeometric function is the generating function of, *inter alia*, the Poisson, binomial, negative binomial, hypergeometric, and Waring distribution. The convergence of Pfaff–Saalschütz to Gauss theorem, applied to the finite population distribution (10), shows that

Proposition 4.2. *For $n \to \infty$, the distribution of ideas' popularity is the Yule–Simon distribution*

$$p(k) = \hat{b}B(k, \hat{b}+1), \tag{12}$$

where B() is the beta function.[8] *The condition $\sum_{k=1}^{\infty} p(k) = 1$ can be verified using Gauss hypergeometric theorem.*

[6] A similar phenomenon was found by Peruani et al. (2007) on the degree distribution of the *fixed* set of nodes, in a growing two-mode network with mixed (random and preferential) attachment, and a high value of the parameter tuning the relative amount of preferential versus random attachment. The model of Evans and Plato (2008), which is a fixed two-mode network with rewiring, can also produce a U-shaped distribution, when the relative amount of preferential vs. random attachment is high.

[7] More general cases involving 5-parameters generalized hypergeometric functions are given in Johnson et al. (2005) and Gutiérrez Jáimez and Rodríguez Avi (1997).

[8] The beta function is defined in terms of the Gamma function: $B(x,y) = \Gamma(x)\Gamma(y)/\Gamma(x+y)$. The Gamma function generalizes the factorial function for non integer values, such that when x is an integer $\Gamma(x+1) = x\Gamma(x) = x!$, but x can also take non-integer values. It relates to the Pochhammer symbol through equation (9).

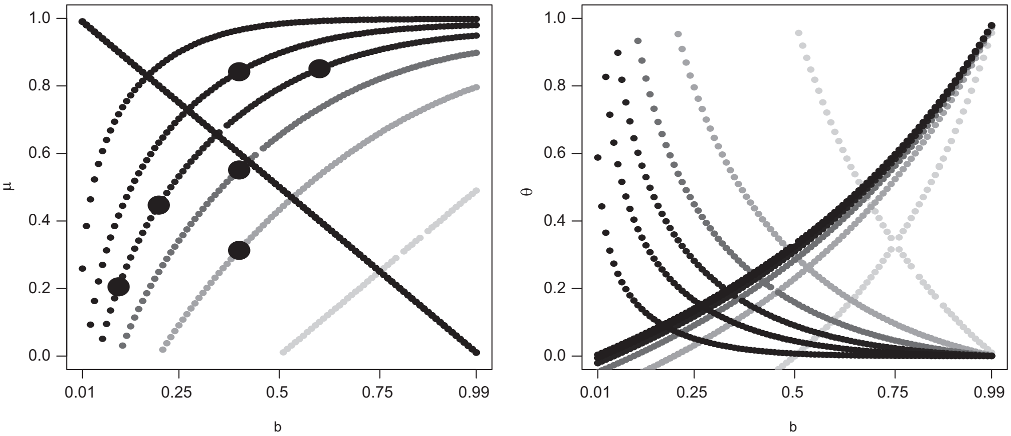

Fig. 3. Left panel: Numerically computed fixed points of Eq. (4) at the steady-state (i.e. solutions of Eq. (13). From the clearest to darkest points, $n=2,5,10,20,50,500$. The six large black dots correspond to the six points of the parameter space used in Fig. 2. Their position with respect to the line $\mu=1-b$ (above or below) determines the shape of the curvature in Fig. 2 (downward or upward). Note that darker dots overlap brighter ones along the line $\mu=1-b$. Right panel: The decreasing curves represent the average overlap θ computed using Eq. (14) (the increasing curves correspond to the fixed point $\mu=1-b$).

Proof. Consider the limit of each term of (10). Assuming $\lim_{n\to\infty}\mu=1$, as will be justified in Section 4.2 (see Fig. 3, left), $\lim_{n\to\infty}p(1)=\hat{b}/(\hat{b}+1)$. Also, $\lim_{n\to\infty}r_1=2+\hat{b}$. Furthermore, $2-n-\hat{b}\mu<r_2<2-n$ and (Slater, 1966, p. 49) $\lim_{n\to\infty}(1-n)_{k-1}/(2-n-\hat{b}\mu)_{k-1}=\lim_{n\to\infty}(1-n)_{k-1}/(2-n)_{k-1}=1$ so $\lim_{n\to\infty}(1-n)_{k-1}/(r_2)_{k-1}=1$. Combining all three limits, $\lim_{n\to\infty}p(k)=(\hat{b}/(\hat{b}+1))(1)_{k-1}/(2+\hat{b})_{k-1}=\hat{b}/(\hat{b}+1)(1+\hat{b})B(k,1+\hat{b})$ which simplifies to (12) □

It should be emphasized that this result is not a steady-state result. The steady-state result for all values of n is given in Proposition 4.1. However, when n is very large and t is not large enough for finite-size effects to be observed (no idea has had enough time to diffuse to the whole population), the Yule–Simon gives a good approximation of the observed distribution.[9]

A last remark on the distribution (10) is its relation to the beta distribution. In the mean field-deterministic-continuous approximation of the stochastic process, the variable k/n follows a distribution proportional to $(k/n)^{-1-\hat{b}\mu}(1-k/n)^{-1+\hat{b}\mu}$ (see Appendix A). However, the support is on $[1/n,1]$ instead of $[0,1]$ for the classical beta distribution, and the restriction on the parameters in the beta (both parameters must be positive) does not hold. The mean-field deterministic approximation is also useful to see that the (expected) diffusion is logistic (Eq. (A.2)).

The distribution (10) is not fully closed form, in the sense that the term μ appears in it, while also depending on it. I now turn to determining the steady-state value of μ.

4.2. Properties of the partition factor

Practically, to compute the predicted steady-state distribution, the value $\mu(b,n)$ is needed. This value can be recorded from the simulations, using either (4) or (6), which are equal by Proposition 3.1. However, it is also possible to compute, prior to the simulations, the tables of μ at its steady-state (so that (10) is genuinely closed-form), for all values of b and n. The steady-state value of μ attained by the stochastic system turns out to be unique, even though the self-consistency fixed point equation studied below admits a second fixed point in the interval of interest. The second fixed point is $\mu=1-b$ for all values of n and can be formally proven.[10] As already mentioned, this fixed point separates the two regions of the parameter space for which there exists or not an upward curvature in the steady state distribution (10).

In the general case the objective is to solve Eq. (4) for μ with $\langle k^2\rangle$ taken at its steady state value. The steady-state value of $\langle k^2\rangle$ and other moments are readily determined

[9] Note that the Yule–Simon distribution can be obtained by taking the $n\to\infty$ limit and putting $\mu=1$ directly in the attachment kernel equation (5). In this case the master equation is the same as in Simon (1955).

[10] Upon substituting $\mu=1-b$, which cancels b, one obtains the surprising one-parameter generalized hypergeometric function identity

$$ {}_3F_2\left[\{2,2,1-n\},\{\tfrac{1}{2}\left(4-n+\sqrt{n(n+4)}\right),\tfrac{1}{2}\left(4-n-\sqrt{n(n+4)}\right)\},1\right]=2n-1. $$

It can be proven using the computer implementation of Gosper's (1978) algorithm by Paule and Schorn (1995). On this topic, see Petkovšek et al. (1996).

Proposition 4.3. *The moments of the popularity distribution are (for* $r \geq 2$*)*

$$\langle k^r \rangle = p_1 \, {}_{r+1}F_r[\{2,2,2,\ldots,1-n\},\{1,1,\ldots,r_1,r_2\},\{1\}].$$

Proof. Each successive term is found by multiplying by $k=(2)_{k-1}/(1)_{k-1}$. □

Inserting the steady-state value of $\langle k^2 \rangle$ and w_t in Eq. (4) gives the fixed point equation

$$\mu = 1 - \frac{b}{n} p_1 \, {}_3F_2[\{2,2,1-n\},\{r_1,4-n-r_1\},1]. \tag{13}$$

This equation is solved numerically in the region of interest ($b \in]0,1[$). I computed values of f_μ (the RHS of the equation) for 99 values of b and a few values of n, and then obtained the fixed points by studying at which points $\mu - f_\mu$ changes sign. The results are reproduced in Fig. 3 where one can see, abstracting from the $\mu = 1-b$ line, that μ is monotonically increasing and concave in b and n. When the population is large, or when innovation is high, an agent i knows only a small proportion of all ideas, and hence the popularity of the ideas unknown by i is high as compared to the total popularity of all ideas. By Proposition 3.1, this implies a higher μ.

For small values of n, μ can be found explicitly but at considerable computational cost. It involves solving polynomials of the order of n. As it turns out, this polynomial always has a root in $\mu = 1-b$. When $n=2$, Appendix C shows that the other root is

$$\mu(n=2) = -\frac{1}{2} + b.$$

Finally, note that the values of μ recorded directly from the simulations (unreported) are in good agreement with the numerical solution of Eq. (13). However, for values of μ close to 0, a significant departure can be observed, especially for low n. This is due to the fact that in these cases, many innovation events occur because of learning events failing (the chosen friend does not have any original ideas to offer, see footnote 5). So, in these cases, the recorded value of μ is not in very good agreement with the input value of b, chosen in advance of the simulation as a parameter; however, it is in excellent agreement with the effective value of b computed as $b = w_t/t$.

It should be emphasized that the convergence of $\mu(t)$ to a fixed point indicates the self-organization of the system. Self-organization results from the feedback loop between structure and dynamics, which comes from the fact that what is learned depends on what is known/unknown. Because μ is a structural quantity capturing the organization of who knows what (Proposition 3.1), and because it determines who learns what (Eq. (5)), it is a fundamental quantity. Its convergence to a fixed point reveals that at this level of aggregation, the system self-organizes into a stable state. To understand this more intuitively, the next section shows that the average pairwise overlap is related to μ.

4.3. Average overlap

Consider the average overlap between two given agents, defined as the Jaccard index of their knowledge portfolios:

$$\theta_{ii'} = \frac{|N_i \cap N_{i'}|}{|N_i \cup N_{i'}|} = \frac{|N_i \cap N_{i'}|}{|N_i| + |N_{i'}| - |N_i \cap N_{i'}|}.$$

The average over all pairs of agents is

$$\theta \equiv \langle \theta_{ii'} \rangle \approx \frac{\langle |N_i \cap N_{i'}| \rangle}{\langle |N_i| + |N_{i'}| - |N_i \cap N_{i'}| \rangle} = \frac{\langle |N_i \cap N_{i'}| \rangle}{2\langle |N_i| \rangle - \langle |N_i \cap N_{i'}| \rangle}.$$

The first relationship is not exact because the expectation of a ratio is, in general, different from the ratio of expectations. However, pairs are very similar in terms of the sizes of their intersections and unions, so that the distribution of these sizes are very tightly peaked, making the approximation fairly good. Now one can use $\langle |N_i| \rangle = t/n$, Lemma 2 and Eq. (4), to get

Proposition 4.4. *The average overlap between agents is well approximated by*

$$\theta = \frac{1-\mu-1/n}{1+\mu-1/n}. \tag{14}$$

Since μ is monotonically increasing in b, the average overlap θ decreases with innovation and increases with learning. Intuitively, an agent who learns ideas of others gets closer to them, and an agent who invents his own ideas increases his distinctiveness. It can also be seen in Fig. 3 (right panel) that θ is also decreasing in n, because it is harder to maintain a high overlap with everybody when there are many agents.

Since there is a one-to-one mapping between b and μ, Eq. (14) implies a one-to-one mapping between b and θ. Hence, for a given number of agents, the rate of innovation determines the average overlap between two agents' portfolio. If the model is reversed in the sense that agents choose to imitate or innovate so as to have a certain θ^*, then, given n, the effective $b = w_t/E_t$ is uniquely determined. In other words, while assuming a fixed innovation–imitation trade-off produces a certain (self-)organization, assuming a certain self-organization would determine the innovation–imitation trade-off. This of course would depend on the particular assumptions made, and would require further work to be rigorously analyzed.[11] However, this remark helps to emphasize that in this model, if there exists an optimal overlap θ^* there also exists an optimal rate of innovation b^*.

5. A few generalizations

5.1. Social network

The derivation of the distribution (10) was made by assuming a complete social network. Consider an opposite case.

Proposition 5.1. *If the social network is a circle in which agents have one friend on each side, the distribution is geometric (with a slight modification for $p(n)$):*

$$p(k) = b(1-b)^{k-1} \text{ for } k \in [1, n-1],$$
$$p(n) = (1-b)^{n-1}.$$

Proof. See Appendix D.

For other types of social networks, simulation results are reported in Fig. 4. If we stay with a one dimensional circular lattice, as in Proposition 5.1, but with a larger number of neighbors on each side, this creates the possibility for an idea to be known by two neighbors of an agent, and the derivation above becomes inexact. However, this configuration would not happen very often, so that for circle networks with small degree, the distribution stays geometric (panel *e* in Fig. 4). However, when the number of neighbors increases to a maximum, the network becomes complete, so that the degree distribution converges to the one obtained under the complete network assumption (see panel *d*). Note that the important criterium to determine the shape of the popularity distribution is not the average degree of an agent, because the competition among ideas cancels out this effect. For instance, panels *a* and *b* show that even sparse Erdős–Renyi networks give results roughly similar to complete networks.[12] The decisive criterion is the dependence or the independence of the attachment kernel on k_j, that is, the fact that the rate of diffusion of an idea depends or not on its popularity. While relating arbitrary social network structure to the popularity distribution by analytical methods is out of the scope of this paper,[13] it can be argued heuristically that social networks with a relatively high number of short cycles will tend to produce an attachment kernel which is not preferential, whereas networks which are closer to trees, such as Barabási–Albert or Erdős–Rényi networks, will tend to produce attachment kernel which are preferential. This comes from the fact that when there are short cycles in the social network, ideas do not diffuse mostly to agents whose neighbors are ignorant of that idea.

5.2. Differentiated productivity

This section relaxes the unrealistic assumption that conditional on investing one unit of time, agents get as many ideas by learning as by innovating. Instead of learning or creating one single idea, agents now have a fixed productivity. When they innovate, they create λ_P ideas, and when they learn, they learn λ_L ideas (sampling a new neighbor with replacement every time).[14] The attachment kernel is now given by

$$A_t(k) = (1-b)\lambda_L \frac{P(j \in N_{i'} \backslash N_i)}{\sum_j P(j \in N_{i'} \backslash N_i)},$$

where $P(j \in N_{i'} \backslash N_i) = k(n-k)/n(n-1)$ does not change. The productivity of learning does not change the nature of the diffusion process, but simply its speed. The productivity of innovation now determines the total number of ideas, $w_t = b\lambda_P t$, and the total number of edges $E_t = t(b\lambda_P + (1-b)\lambda_L)$. It still holds that $\sum_{j=1}^{w} P(j \in N_{i'} \backslash N_i) = (nE_t - w_t \langle k^2 \rangle)/n(n-1)$ so that

$$A_t(k) = (1-b)\lambda_L \frac{k(n-k)}{nE_t - w_t \langle k^2 \rangle} = \frac{k(n-k)}{(\zeta+1)\mu n t},$$

where μ is still defined by Eq. (4), and the combinatorial interpretation (Proposition 3.1) still holds. The parameter ζ is defined as $\zeta = b\lambda_P/(1-b)\lambda_L$. Note that if we set $\lambda_P = \lambda_L = 1$, we find $\zeta + 1 = \hat{b}$ as it must to recover the attachment kernel (5). The procedure to find the steady-state distribution (Section 4.1) can be followed here as well. The resulting degree distribution simply now balances the rate of innovation $b\lambda_P$ with the rate of learning $(1-b)\lambda_L$ (instead of only b with $1-b$).

[11] Using simulations, I checked the following. At each period, a directed pair of agents is chosen at random. Agent i computes $\theta_{ii'}$ and learns from i' if $\theta_{ii'}$ is less than some predetermined θ^*, otherwise he innovates. As expected, in the simulations explored, pairs of agents converge to $\theta_{ii'} \approx \theta^*$ and the system itself exhibits $w_t/t = b_{effective} \approx b_{theory}$ where b_{theory} is computed by inverting the relationship between θ and b through μ.

[12] There was a probability that these random networks would not be fully connected. I forced connectedness by repeating the algorithm of graph creation until one fully connected was created – so strictly speaking, they are not Erdős–Rényi but random graphs from a (slightly) restricted ensemble.

[13] It should be stressed that the derivations for Proposition 5.1 and for the complete social network make use of the assumption that the pairs are homogeneous, which works very well in these cases, but this may not always be the case.

[14] λ_L must be a small number to ensure that there are enough ideas to be learned. See footnote 5.

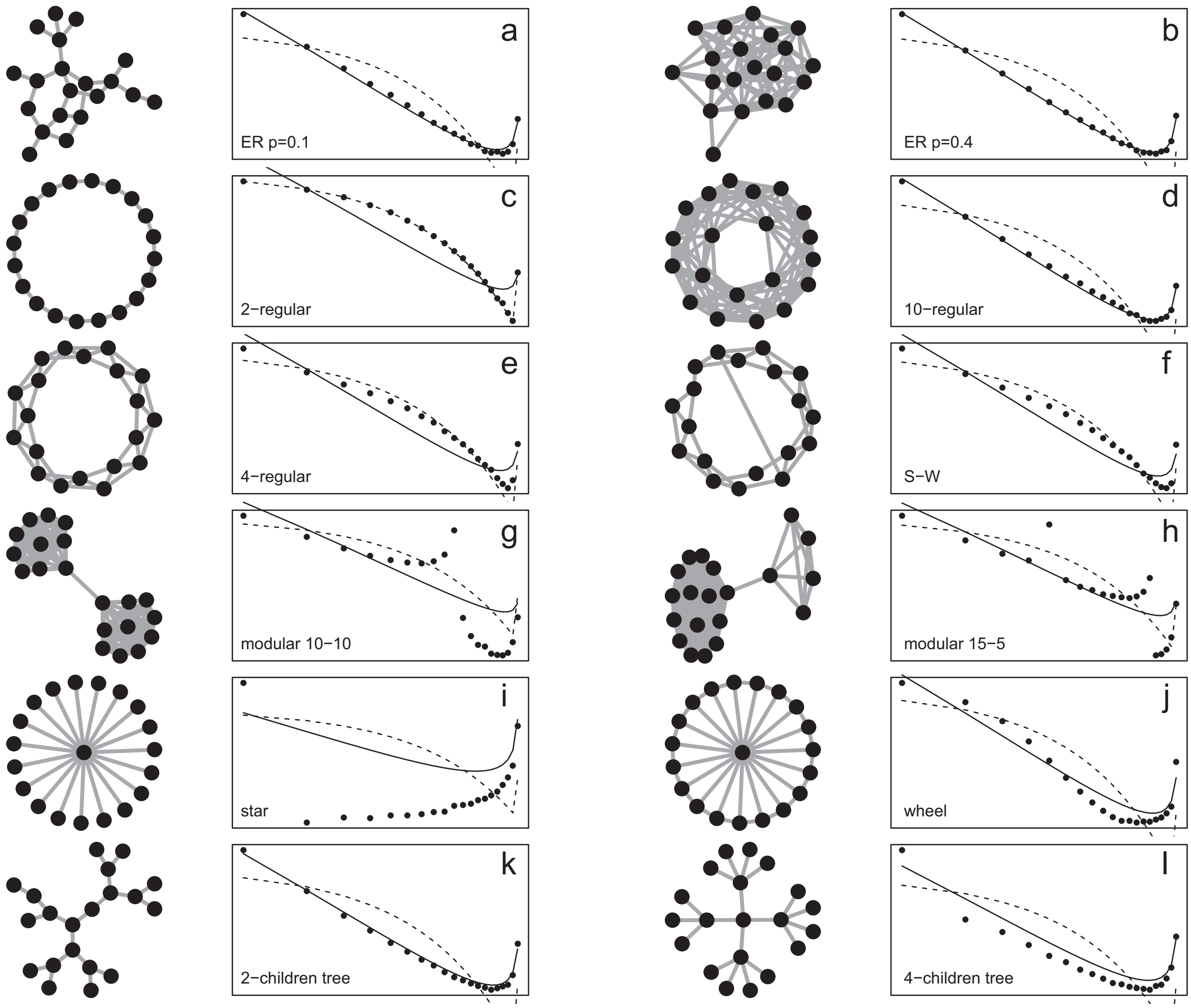

Fig. 4. Simulations for 10^6 periods, using different social networks, $n=20$ and $b=0.2$. The solid line is the theoretical result for a complete social network, as in Fig. 2. The dashed line is the theoretical result for one-dimensional 2-regular circular lattice (Proposition 5.1). The first line of the panel shows two (connected) Erdős–Rényi ($p=0.1, 0.4$). The second line shows two q-regular circular one dimensional lattice ($q=2, 10$). The third line shows two 4-regular one dimensional circular lattice, with one rewired edge for the right one. The fourth line shows two modular networks, constructed by linking two complete subgraphs. The fifth line shows a star network, and a combination of a star and a one dimensional 2-regular circular lattice. The last line shows two regular trees, one with two children, and one with four children (some nodes have less children, due to the requirement that $n=20$).

In the limit of an infinite population, the exponent of the Yule–Simon was $2+b/(1-b)$, and with productivity parameters it can be shown that it is $2+\zeta$. This highlights that the original and productivity-augmented models can really be thought of as one parameter (ζ) models.

6. Conclusion

The importance of innovation and knowledge diffusion in economic systems is widely recognized. Likewise, the literature has emphasized the role of interactions and self-reinforcing dynamics in shaping the structure and dynamics of economies. This context calls for a fundamental understanding of the self-organization of knowledge economies: considering interacting agents who innovate new-to-the-world ideas and imitate existing ideas, what can we say about the likely long term structure of who knows what?

In this paper, I have characterized a parsimonious model of knowledge diffusion and growth. In the model, learning ideas of friends implies self-reinforcing but bounded diffusion, leading to a logistic diffusion curve. Together with a continuous arrival of new ideas, a stable organization emerges in terms of the distribution of ideas' popularity. In general, because who knows what determines who learns what, there is a feedback loop between the structure and the dynamics of the system. In the model studied, this feedback loop leads to a stable state which is thus “self-organized”. Moreover, it is shown that under most circumstances, the distribution of ideas' popularity is close to a power law: most ideas are known by only a few agents, who just discovered them, and only a few ideas have diffused completely. A higher rate of diffusion, relative to the rate of innovation, implies a heavier right tail of the popularity distribution (more diffusion implies more very well-known ideas). Moreover, the structure of the social network on which ideas' diffusion takes place matters in a non-trivial way. Heuristically, networks with short cycles, as opposed to tree-like networks, prevent ideas from diffusing as fast as they could, leading to less skewed distribution of ideas' popularity.

The main consequence of these results is that in a society which facilitates relatively more diffusion than innovation (which implies a high productivity of learning, λ_L, and a low relative probability of innovation, b, if the choice of innovation/imitation depends on the relative returns to each activity), we should expect the distribution of ideas' popularity to be very skewed and the average overlap to be very high. On the other hand, in a society which favors the emergence of genuinely new ideas, we should expect the distribution of ideas' popularity to be less unequal, and the average overlap to be lower. Intuitively, societies in which diffusion is much more intense than innovation will tend to be more cohesive, in the sense that agents will tend to have similar ideas.

Finally, the model contributes to the ongoing research agenda on the evolution of networks. Since it is a nonlinear (logistic) extension of the widely used model of Simon (1955), and given the large number of phenomena across which power laws and logistic growth are observed, it might be of interest beyond the context of knowledge systems.

Acknowledgments

I thank Robin Cowan, Luis Lafuerza, Daniel Opolot, Giorgio Triulzi, two anonymous referees and audiences at Eindhoven, Oxford, Santa-Fe and the INSNA and WEHIA conferences. Financial support from Maastricht Graduate School of Business and Economics is gratefully acknowledged. All errors are mine.

Appendix A. Distribution of ideas' popularity: mean-field continuous deterministic approximation

Consider that each idea j diffuses deterministically and assume that time is continuous. Using (5),

$$\frac{dk_j(t)}{dt} = \frac{k_j(t)(n-k_j(t))}{\hat{b}\mu n t}. \tag{A.1}$$

This is a first-order ordinary differential equation. It looks similar to Verhulst's equation of population growth, except that it has non-constant coefficients since t appears on the RHS. It is nonlinear, but it is a Bernoulli equation so it can be linearized and integrated. The simplest is probably to separate variables to obtain

$$\hat{b}\mu n \int \frac{1}{k(n-k)} dk = \int \frac{1}{t} dt$$

$$\hat{b}\mu \left[\log\left(\frac{k}{k-n}\right) + C_1 \right] = \log(t) + C_2$$

$$k_j(t) = n/(1 - Ct^{-1/\hat{b}\mu}),$$

where C is an arbitrary constant. Using the initial condition $k_j(t_j) = 1$, it follows that $C = (n-1)/(t_j^{-1/\hat{b}\mu})$, and therefore the solution of (A.1) is

$$k_j(t) = n\left[1 + (n-1)\left(\frac{t_j}{t}\right)^{(1-b)/\mu}\right]^{-1}. \tag{A.2}$$

Note that Eq. (A.2) has a logistic form with an inflexion point at some $k > 0$ if $\mu < 1-b$. The continuous distribution is computed thus (using Eq. (A.2)):

$$\begin{aligned}\Pr(k_j \le k) &= \Pr\left(n\left[1 + (n-1)\left(\frac{t_j}{t}\right)^{(1-b)/\mu}\right]^{-1} \le k\right) \\ &= 1 - \Pr\left(t_j \le \left(\frac{k_j - k_j n}{k_j - n}\right)^{\hat{b}\mu} t\right).\end{aligned}$$

Since the t_j's are uniformly distributed[15] their probability mass function is $\Pr(t_j = Y) = 1/t$ for Y from 1 to t, so $\Pr(t_j \le Y) = \sum_1^Y 1/t = Y/t$. This leads to

[15] Contrary to one-mode scale free network models, this is not exactly true, since there is not one new idea per period, but only one at each period with probability b. The uniform distribution is, nevertheless, an appropriate approximation since the t_js of many independent realizations of the stochastic process are uniformly distributed over $[1, t]$.

$$\Pr(k_j \le k) = 1 - \left(\frac{k_j - k_j n}{k_j - n}\right)^{\hat{b}\mu}.$$

Using $p(k) = dPr(k_j \le k)/dk$ gives

$$p(k) = \hat{b}\mu n(n-1)^{-\hat{b}\mu}(n-k)^{-1+\hat{b}\mu}k^{-1-\hat{b}\mu}. \tag{A.3}$$

One can check that this is a proper distribution function, $\int_1^n p(k)\,dk = 1$. This distribution has the shape of a particular beta distribution. Making the change of variable $x = k/n$, we have

$$p(x) \propto (1-x)^{-1+\hat{b}\mu}x^{-1-\hat{b}\mu},$$

which is almost the definition of a beta distribution beta (α,β) with $\alpha = -\hat{b}\mu$ and $\beta = \hat{b}\mu$. However, negative parameters are not allowed in the definition of the beta distribution. Moreover, the factor of proportionality is different from that of the beta distribution because the support is different. This distribution has to have a strictly positive support, because the integral diverges at 0.

Appendix B. Distribution of agents' number of ideas known

The number of ideas known by a r.c. agent has a binomial distribution. To see this, note that to "know" k_a ideas at time t, a r.c. agent needs to have been chosen exactly k_a times, and not chosen exactly $(t-k_a)$ times. Thus it follows that the distribution of agents' number of ideas known is the binomial distribution:

$$p_t(k_a) = \binom{t}{k_a}\left(\frac{1}{n}\right)^{k_a}\left(1-\frac{1}{n}\right)^{t-k_a}.$$

Appendix C. Exact solution of the fixed point equation for $n=2$

Written explicitly the fixed point equation (13) becomes

$$\mu = 1 - \frac{b}{n}p(1)\sum_{k=1}^{n} k^2 \frac{(1)_{k-1}(1-n)_{k-1}}{(r_1)_{k-1}(4-n-r_1)_{k-1}}. \tag{C.1}$$

For $n=2$, the series has only two terms and is $1+2/\hat{b}\mu$. The fraction of ideas known one time is $(p_1|n=2) = 2\hat{b}\mu/2\hat{b}\mu+1$. Substituting into (C.1) and simplifying lead to the quadratic equation in μ:

$$-2\hat{b}\mu^2 + \hat{b}\mu + 1 - 2b = 0,$$

which has solutions

$$\mu = \{1-b, -1/2+b\}. \tag{C.2}$$

This result can also be derived using the inclusion/exclusion formula or by writing the dynamic process for μ_i.

Appendix D. Distribution of ideas' popularity when the social network is a circle

Consider a network in which agents are placed around a circle and have only one friend on each side. Because ideas diffuse face to face, the number of social network (directed) pairs with $j \in N_i \cap N_{i'}$ is simply $2(k_j-1)$. It is also easy to see that there are only two directed pairs such that $j \in N_{i'} \backslash N_i$. In total, there are $2n$ directed pairs. Thus $\Pr(j \in N_{i'} \backslash N_i) = 2/2n$, so that $|N_{i'} \backslash N_i| = w/n$. The attachment kernel is then $A_t(k_j) = (1-b)/bt$, and the master equation for the steady state becomes $p(k) = (1-b)p(k-1)$. The first term is found to be $p_1 = b$, hence iterating the master equation gives the geometric distribution:

$$p(k) = b(1-b)^{k-1}.$$

However, when an idea is known n times, it cannot diffuse more. There are no bias as long as $k \le n-1$, but for $k > n$ it must be that $p(k) = 0$. For $k=n$ the master equation becomes

$$p(n) = \frac{(1-b)}{b}p(n-1) - 0$$

$$p(n) = \frac{(1-b)}{b}b(1-b)^{n-2} = (1-b)^{n-1}.$$

The key point in the derivation above is that $\Pr(j \in N_{i'} \backslash N_i)$ is independent of k_j. As long as this is the case, the same distribution will be obtained, because of the normalization by the sum (the competition among ideas).

References

Arthur, W.B., 1989. Competing technologies, increasing returns, and lock-in by historical events. Econ. J. 99 (394), 116–131.
Barabási, A., Albert, R., 1999. Emergence of scaling in random networks. Science 286 (5439), 509.
Beguerisse-Díaz, M., Porter, M.A., Onnela, J.-P., 2010. Competition for popularity in bipartite networks. Chaos: An Interdisciplinary Journal of Nonlinear Science 20 (4).
Börner, K., Maru, J., Goldstone, R., 2004. The simultaneous evolution of author and paper networks. Proc. Natl. Acad. Sci. USA 101 (Suppl 1), 5266–5273.
Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. J. Econ. Geogr. 9 (4), 439–468.
Cohen, W., Levinthal, D., 1989. Innovation and learning: the two faces of R & D. Econ. J. 99 (397), 569–596.
Cowan, R., Jonard, N., 2004. Network structure and the diffusion of knowledge. J. Econ. Dyn. Control 28 (8), 1557–1575.
Cowan, R., Jonard, N., 2009. Knowledge portfolios and the organization of innovation networks. Acad. Manag. Rev. 34 (2), 320–342.
de Solla Price, D., 1976. A general theory of bibliometric and other cumulative advantage processes. J. Am. Soc. Inf. Sci. 27 (5), 292–306.
Evans, T., Plato, A., 2008. Network rewiring models. Netw. Heterog. Media 3 (2), 221–238.
Geroski, P.A., 2000. Models of technology diffusion. Res. Policy 29 (4), 603–625.
Gosper, R.W., 1978. Decision procedure for indefinite hypergeometric summation. Proc. Natl. Acad. Sci. USA 75 (1), 40–42.
Gutiérrez Jáimez, R., Rodríguez Avi, J., 1997. Family of Pearson discrete distributions generated by the univariate hypergeometric function ${}_3F_2\,(\alpha_1, \alpha_2, \alpha_3; \gamma_1, \gamma_2; \lambda)$. Appl. Stoch. Models Data Anal. 13 (2), 115–125.
Ijiri, Y., Simon, H.A., 1977. Skew Distributions and the Sizes of Business Firms. North-Holland Publishing Company, New York.
Jackson, M., Rogers, B., 2007. Meeting strangers and friends of friends: how random are social networks?. Am. Econ. Rev. 97 (3), 890–915.
Johnson, N.L., Kemp, A.W., Kotz, S., 2005. Univariate Discrete Distributions, vol. 444. John Wiley & Sons, New York.
Jovanovic, B., Rob, R., 1989. The growth and diffusion of knowledge. Rev. Econ. Stud. 56 (4), 569–582.
Kirman, A., 1993. Ants, rationality, and recruitment. Q. J. Econ. 108 (1), 137–156.
König, M., Lorenz, J., Zilibotti, F., 2012. Innovation Vs Imitation and the Evolution of Productivity Distributions. CEPR Discussion Paper 8843.
Krapivsky, P., Redner, S., Leyvraz, F., 2000. Connectivity of growing random networks. Phys. Rev. Lett. 85 (21), 4629–4632.
Lafond, F., 2014. Knowledge diffusion and the structure of citation networks. In: The Evolution of Knowledge Systems. UNU-MERIT (Ph.D. thesis). Maastricht University Press, pp. 73–89.
Liu, C., Yeung, C., Zhang, Z., 2011. Self-organization in social tagging systems. Phys. Rev. E 83 (6), 066104.
Lucas, R.E., 1988. On the mechanics of economic development. J. Monet. Econ. 22 (1), 3–42.
Lucas, R.E., Moll, B., 2014. Knowledge growth and the allocation of time. J. Polit. Econ. 122 (1), 1–51.
Mansfield, E., 1961. Technical change and the rate of imitation. Econometrica 29 (4), 741–766.
Paule, P., Schorn, M., 1995. A Mathematica version of Zeilberger's algorithm for proving binomial coefficient identities. J. Symb. Comput. 20 (5–6), 673–698.
Peruani, F., Choudhury, M., Mukherjee, A., Ganguly, N., 2007. Emergence of a non-scaling degree distribution in bipartite networks: a numerical and analytical study. Europhys. Lett. 79, 28001.
Petkovšek, M., Wilf, H., Zeilberger, D., 1996. A=B. Wellesley, MA, ⟨http://www.math.rutgers.edu/~zeilberg/AeqB.pdf⟩.
Ramasco, J., Dorogovtsev, S., Pastor-Satorras, R., 2004. Self-organization of collaboration networks. Phys. Rev. E 70 (3), 036106.
Romer, P.M., 1990. Endogenous technological change. J. Polit. Econ. 98 (5), S71–S102.
Roth, C., Cointet, J., 2010. Social and semantic coevolution in knowledge networks. Soc. Netw. 32 (1), 16–29.
Schultz, T.W., 1961. Investment in human capital. Am. Econ. Rev. 51 (1), 1–17.
Schumpeter, J.A., 1934. The Theory of Economic Development: An Inquiry into Profits, Capital, Credit, Interest, and the Business Cycle. Harvard University Press, Cambridge.
Sen, A., 1999. Development as Freedom. Alfred Knopf, New York.
Simon, H., 1955. On a class of skew distribution functions. Biometrika 42 (3/4), 425–440.
Slater, L.J., 1966. Generalized Hypergeometric Functions. Cambridge University Press, Cambridge.
Solow, R.M., 1957. Technical change and the aggregate production function. Rev. Econ. Stat. 39 (3), 312–320.
Vázquez, A., 2003. Growing network with local rules: preferential attachment, clustering hierarchy, and degree correlations. Phys. Rev. E 67 (5), 056104.
Young, H.P., 2009. Innovation diffusion in heterogeneous populations: contagion, social influence, and social learning. Am. Econ. Rev. 99 (5), 1899–1924.
Zeng, A., Yeung, C., Shang, M., Zhang, Y., 2012. The reinforcing influence of recommendations on global diversification. Europhys. Lett. 97 (1), 18005.