Working Paper Series

#2013-040

Self-organization of knowledge economies

François Lafond

Maastricht Economic and social Research institute on Innovation and Technology (UNU-MERIT)
email: info@merit.unu.edu | website: http://www.merit.unu.edu

Maastricht Graduate School of Governance (MGSoG)
email: info-governance@maastrichtuniversity.nl | website: http://mgsog.merit.unu.edu

Keizer Karelplein 19, 6211 TC Maastricht, The Netherlands
Tel: (31) (43) 388 4400, Fax: (31) (43) 388 4499

UNU-MERIT Working Papers

ISSN 1871-9872

Maastricht Economic and social Research Institute on Innovation and Technology, UNU-MERIT

Maastricht Graduate School of Governance MGSoG

UNU-MERIT Working Papers intend to disseminate preliminary results of research carried out at UNU-MERIT and MGSoG to stimulate discussion on the issues raised.

Self-organization of knowledge economies*

François Lafond †

August 2, 2013

Abstract

Suppose that homogenous 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 downward curvature. For infinite population it converges to the Yule-Simon distribution. The power law is steeper when innovation is high. Diffusion follows S-shaped curves.

Keywords: innovation, diffusion, two-mode networks, cumulative advantage, quadratic attachment kernel, power law, Yule-Simon distribution, generalized hypergeometric distribution.

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 2001). At the firm level, innovation is the source of competitive advantage and profits (Schumpeter 1934). At the country level, technical 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

*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. Here I give additional results and improved proofs for the basic model, and I do not include a citation network. I thank Robin Cowan, Luis Lafuerza, Daniel Opolot, Giorgio Triulzi, and the participants to ECIS seminar and INSNA conference. Financial support from METEOR is gratefully acknowledged. All errors are mine.

†Address: UNU-MERIT, Keizer Karelplein 19, 6211 TC, Maastricht, The Netherlands. E-mail: *lafond@merit.unu.edu*.

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 & Jonard 2004). My intention here is similar to general equilibrium economics. Under which conditions there exists a form of balanced state in which agents produce and exchange ideas? Is this state unique? I describe a simple stochastic process of innovation and imitation which has a unique mean-field steady-state. The main assumptions and results are 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. I assume that ideas are indistinguishable, except for their age and who knows them.*

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. (the friendship network is assumed to be complete).*

Assumption 3 (Limited attention and innovation/imitation trade-off)**.** *Homogenous 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.*[1]

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.*[2] *However, diffusion is constrained by the population size, as in logistic diffusion models. Thus, ideas' diffusion is S-shaped.*

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 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.*

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

[2] 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.

In the model, new ideas arrive regularly, and each idea is eventually known by everybody. The competition of ideas for the attention of agent i can be characterized by computing which ideas are known by a friend of i and unknown to i. This could never be completely stable nor exactly equal across pairs. However, at the mean-field level, there exists a fixed point, self-consistency equation which allows to derive a steady-state that is unique.

The proposed model is an extension of Simon's (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's heterogeneity, and consider a fixed number of agents. I still want to have a growing number of ideas (assumption 1), consistent with reality, but also wish to allow agents to learn ideas of/from others (assumption 2). Since I contend that attention is limited (assumption 3), I assume that at each period, a randomly 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. 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 S-shaped. This imposes to modify Simon's master equation for the degree distribution, 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 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 & 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. The literature on diffusion is huge (Geroski 2000), but 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. 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 S-Shaped, 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, this exponential growth is constrained by the 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 of ideas' popularity.[3] 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 & Cointet 2010) and simulation modelling (Börner et al. 2004). Cowan & Jonard (2009) study a closely related system, where firms form an alliance network

[3] For models of knowledge growth and diffusion which do not involve networks, see e.g. Jovanovic & Rob (1989), König et al. (2012) and Lucas & Moll (2013). The model presented here is complementary, because these models are more elaborated in terms of agent's choice 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 & Moll (2013).

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 wold properties and skewed degree distribution.

The model is also closely related to models of network growth based on copying (Vázquez 2003). For instance, in Jackson & 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 more chances he has to be found through a friend. Likewise, in a two-mode network, search can generates cumulative advantage and, ultimately, fat tail distribution of popularity. This was clearly demonstrated by Evans & Plato (2008), who consider a fixed set of agents and a fixed set of artifacts. Agents are linked to one and only one artifact, and, when they are chosen, connect to another artifact 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 artifact, and the distribution of artifacts popularity is studied. Their model applies for instance in anthropology where one is interested in the transmission of cultural artifacts. The model proposed below also applies to this context, but assumes that new artifacts appear over time, and that actors accumulate artifacts over time.

Another closely related model was studied by Ramasco et al. (2004). As Simon, they consider 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. For instance, Beguerisse-Diaz 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 artifact 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 presented below 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).

Following Simon's own applications of his model, notably to the size distribution of firms (Ijiri & Simon 1977), there has been a large literature impossible to review here. In an influential contribution, 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 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 2012), using the model described below, and assuming that (an infinite number of) agents cite papers chosen u.a.r among the papers that they have previously learned or written, the predicted citation distribution is a shifted power law. Hence, in the model below, the arrival of new ideas may be thought of as emerging from the recombination of existing ideas held by individual inventors.

3 The model

Assuming an infinite population and deterministic diffusion, the model can be summarized as follows. 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 equally count as function of 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 starting from the cumulative distribution function

$$prob(k_j \leq k) = p\left(\left(\frac{t}{t_j}\right)^{1-b} \leq k\right) = 1 - p\left(t_j \leq tk^{-1/(1-b)}\right)$$

Assume that ideas arrived sequentially, in such a way that the t_j's are uniformly distributed i.e. $prob(t_j = Y) = 1/t$ for Y from 1 to t, so $prob(t_j \leq Y) = \sum_1^Y \frac{1}{t} = \frac{Y}{t}$. This leads to $p(k_j \leq k) = 1 - k^{\frac{-1}{1-b}}$. Apply $p(k) = \frac{dp(k_j \leq k)}{dk}$ 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 $\gamma = 2 + \frac{b}{1-b}$ 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 an S-shaped diffusion pattern (see equation 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 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 actor-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.[4]) agent ($w_1 = 1$ and $E_1 = 1$). Then at each time period, the following algorithm is applied (where random always means uniformly at random):

I/ pick an agent i at random.

[4]Throughout the idea, "random" refers to a uniform distribution

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).

3.2 Preliminary results

Consider a matrix Q which has a fixed number of rows (n agents) and a number of columns that depends 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 iff 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:*

[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 happen 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 figure 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 (equation 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 assumption of inelastic supply of (cognitive) labor. It is possible for individual agents to imitate faster than they innovate, because one newly created idea can be imitated $(n - 1)$ times.

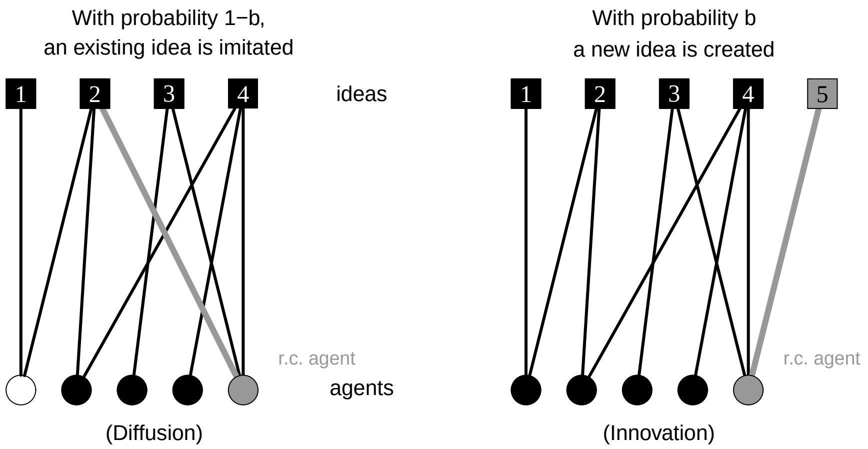

Figure 1: **Schematic description of the model**. At each time step, one and only one of the two events represented above happens. In both cases, a link is added. The main focus here is on the degree distribution of the top nodes (ideas' popularity), $p(k)$. The degree distribution of the bottom nodes is discussed in appendix B but is purposefully uninteresting (agents are homogenous so it is binomial). On the left panel, the r.c. agent 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. actor 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 doesn't know himself— in the example above he turned out to choose the second 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.

Proof.

$$D_t = \frac{E_t}{nw_t} \approx \frac{1}{nb}$$

Hence, 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). 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 doesn't 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*

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

Proof. Observe that the sum over all pairs of $|N_{i'} \cap N_i|$ is simply the total number of "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, we obtain the sum. 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 □

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 I have excluded all structural 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 which conditions idea j will 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 must be such that $j \in \{N_{i'} \backslash N_i\}$. 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 be chosen, that is

$$A_t(k_j) := prob(k_j(t+1) = k_j(t) + 1) \; ; \; \sum_{j=1}^{w_t} A_t(k_j) = 1 - b$$

$$A_t(k_j) = (1-b) \frac{1}{\binom{n}{2}} \sum_{i<i'} \frac{P(j \in \{N_{i'} \backslash N_i\})}{|N'_i \backslash N_i|}$$

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

In the equation above, all pairs (i, i') have an equal chance of being chosen, and pairs are treated symmetrically. The probability that $j \in \{N_{i'} \backslash N_i\}$ can be computed as follows. There are $\binom{n}{2}$ pairs in the system. Idea j has degree k_j so there are $\binom{k_j}{2}$ pairs where $j \in \{N_i \cap N_{i'}\}$

and $\binom{n-k}{2}$ where $j \notin \{N_i \cup N_{i'}\}$. The rest of the pairs are either such that $j \in \{N'_i \backslash N_i\}$ or such that $j \in \{N_i \backslash N_{i'}\}$. If we treat N_i and $N_{i'}$ symmetrically, we find that the number of pairs such that $j \in \{N'_i \backslash N_i\}$ is equal to $\frac{\binom{n}{2}-\binom{k_j}{2}-\binom{n-k_j}{2}}{2} = \frac{k_j(n-k_j)}{2}$. Hence, a r.c. pair will exhibit $j \in \{N'_i \backslash N_i\}$ with probability $\frac{k_j(n-k_j)}{2}/\binom{n}{2} = \frac{k_j(n-k_j)}{n(n-1)}$. Using equation 2, we readily determine $|N'_i \backslash N_i| = \frac{\mu t}{n-1}$. where μ is defined as

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

where I omit the time subscript in $\langle k^2 \rangle = \sum_{j=1}^{w_t}[k_j(t)]^2$. Thus, equation 3 becomes

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

The condition $\sum_{j=1}^{w_t} A_t(k_j) = 1 - b$ is the same equation as the definition of μ (equation 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 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 equation 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.*

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

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

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

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

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

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,

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

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, equation 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.

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 equation 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 k^{th} 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 equation 5

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

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

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

Equation 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 $(-1)(i-u_1)(i-u_2)$ where $\{u_1, u_2\} = \frac{1}{2}\left(2 - n \pm \sqrt{n(n+4\hat{b}\mu)}\right)$. Now consider the definition of the Pochhammer symbol

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

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

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

From Slater (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

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

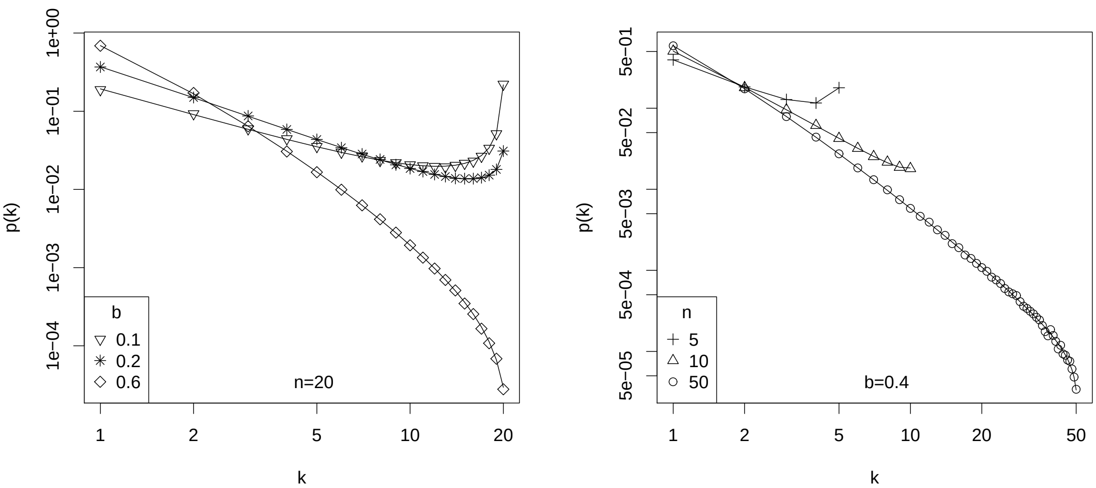

Figure 2: **Distribution of ideas' popularity.** For each of six configurations of parameters, the model is run once for 3×10^6 periods (only 10^6 when $n = 5, n = 10$). In the left panel, $n = 20$ and the effect of b is studied. In the right panel, $b = 0.4$ and the effect of n is studied. The plain lines are the theoretical results, computed using equation 10 and values of μ computed using the fixed point equation (14). These six points of the parameter space are marked in figure 3. When a point in figure 3 is in the lower left half of the (μ, b) plane, the corresponding curve in the figures above exhibit an upward curvature, otherwise it exhibits a downward curvature.

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

The probability mass function (10) is plotted against simulations in figure 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 $k^* = \frac{1}{2}(1 + n(1 + \hat{b}\mu))$ and the point at which $p(k^* + 1) = p(k^*)$ is given by $k^* = \frac{1}{2}(-1 + n(1 + \hat{b}\mu))$ so that we may take $k^* = \frac{1}{2}n(1 + \hat{b}\mu)$. 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 figure 3 (see section 4.2), which 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 (${}_3F_2[]$) 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 Saalschutzian. Hence, the Pfaff-Saalschutz 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-Saalschutzian. 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].

$$G(z) = \sum_{k=1}^{n} p(k) z^k = \frac{{}_3F_2[\{1, 1, 1-n\}, \{r_1, r_2\}, z]}{{}_3F_2[\{1, 1, 1-n\}, \{r_1, r_2\}, 1]} \tag{12}$$

This class is interesting because there exists a deep connection between Pfaff-Saalschutz and Gauss hypergeometric theorem, and Gauss hypergeometric function is the generating

[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 & Plato (2008), which is a fixed two-mode network with rewiring, can also produce a U-shaped distribution, when the relative amount of preferential v.s. random attachment is high.

[7] More general cases involving 5-parameters generalized hypergeometric functions are given in Johnson et al. (2005) and Gutirrez Jimez & Rodrguez Avi (1997).

function of, inter alia, the Poisson, binomial, negative binomial, hypergeometric, and Waring distribution. In fact each theorem can be obtained starting from the other (Slater 1966, p. 48-49). The convergence of Saalschutz 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*

$$p(k) = \hat{b}B(k, \hat{b}+1) \tag{13}$$

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

Proof. Consider the limit of each term of (10). Assuming $\lim_{n\to\infty} \mu = 1$, as will be justified in section 4.2, $\lim_{n\to\infty} p(1) = \frac{\hat{b}}{\hat{b}+1}$. Also, $\lim_{n\to\infty} r_1 = 2 + \hat{b}$. Furthermore, $\lim_{n\to\infty} \frac{1-n}{r_2} = 1$ so $\lim_{n\to\infty} \frac{(1-n)_{k-1}}{(r_2)_{k-1}} = 1$. Combining all three limits, $\lim_{n\to\infty} p(k) = p(1)\frac{(1)_{k-1}}{(2+\hat{b})_{k-1}} = p(1)(1+\hat{b})B(k, 1+\hat{b})$ which simplifies to (13) □

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 $(\frac{k}{n})^{-1-\hat{b}\mu}(1-\frac{k}{n})^{-1+\hat{b}\mu}$(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 S-shaped (equation A.2 on page 20).

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 in advance of 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

[8]The Beta function is defined in terms of the Gamma function: $B(x, y) = \frac{\Gamma(x)\Gamma(y)}{\Gamma(x+y)}$. 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.

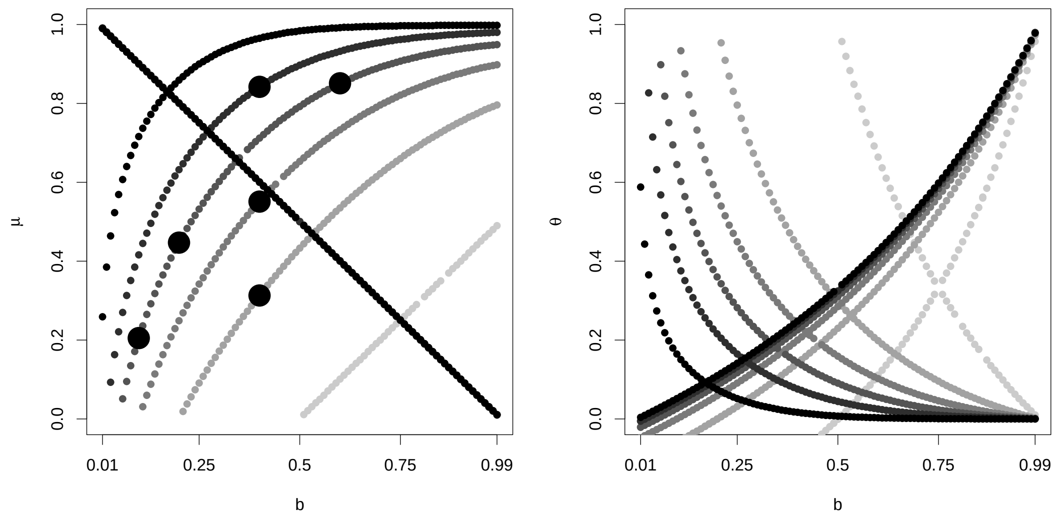

Figure 3: Left panel: Numerically computed fixed points of equation 4 at the steady-state (i.e. solutions of equation 14). From the clearest to darkest points, $n = 2, 5, 10, 20, 50, 500$. The six large black dots corresponds to the six points of the parameter space used in figure 2. Their position with respect to the line $\mu = 1 - b$ (above or below) determines the shape of the curvature in figure 2 (downward or upward). Right panel: The decreasing curves represent the average overlap θ computed using equation 15 (the increasing curves correspond to the fixed point $\mu = 1 - b$).

the interval of interest. The second fixed point is $\mu = 1 - b$ for all values of n. 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 equation 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

Proposition 4.3. *The moments of the popularity distribution are*

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

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

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

$$\mu = 1 - \frac{b}{n} \, p_1 \; {}_3F_2[\{2, 2, 1 - n\}, \{r_1, 4 - n - r_1\}, 1] \tag{14}$$

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 figure 3 where one can see that μ is monotonically increasing and concave in b and n.[9]

[9] Upon substituting $\mu = 1 - b$, which cancels b, one obtains the surprising one-parameter generalized hyper-

For small values of n, μ can be found explicitly but at considerable computational cost. It involves solving polynomials of the order of n. One root will always be $1-b$. When $n=2$, appendix C shows that the other root is

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

4.3 Average overlap

Consider the average overlap between two given agents, defined as follows

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

The average over all pairs of agents is

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

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 we can use $\langle |N_i| \rangle = t/n$, lemma 2 and equation 4, to get

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

$$\theta = \frac{1 - \mu - 1/n}{1 + \mu - 1/n} \tag{15}$$

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 figure 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, equation 15 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.

---

geometric function identity

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

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

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 $p(n)$)*

$$\begin{aligned} p(k) &= b(1-b)^{k-1} \qquad\qquad \textit{for } k \in [1, n-1] \\ p(n) &= (1-b)^{n-1} \end{aligned}$$

Proof. Appendix D

If we allow 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. However, when the number of neighbors increases to a maximum, the network becomes complete, as assumed previously. Note that the important criteria 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, simulations show that sparse Erdős-Renyi networks give results roughly similar to complete networks. The decisive criteria is the dependence or 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.

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 than 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)[10]. 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 $P(j \in N_{i'} \backslash N_i) = \frac{k(n-k)}{n(n-1)}$ does not change. The productivity of learning does not change the nature of the diffusion process, but simply its speed. The productivity of innovation

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

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.$\sum_{j=1}^{w} P(j \in N_{i'} \backslash N_i) = \frac{nE_t - w_t\langle k^2 \rangle}{n(n-1)}$ 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 equation 4, and the combinatorial interpretation (proposition 3.1) still holds. The parameter ζ is defined as $\zeta = \frac{b\lambda_P}{(1-b)\lambda_L}$. 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 $1-b$). In the limit of an infinite population, the exponent of the Yule-Simon was $2 + \frac{b}{1-b}$, 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

A parsimonious model of knowledge growth and diffusion was presented. It gives rises to a stable organization, in terms of the distribution of ideas' popularity. I refrained from extending the model by adding extra assumptions, as I believe that such additional assumptions shall be guided by the particular theoretical problem and/or empirical data one wishes to explore.

At this level of generality, one can conclude that in a society which facilitates relatively more diffusion than innovation (which implies a high λ_L, and a low 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 fall faster, and the average overlap to be lower. If we think that pairs of agents must have a given overlap in equilibrium, then for a given social network the relative rate of innovation and imitation is uniquely determined.

Besides applications related to the organization of knowledge systems, the model might be of interest in the numerous domains where Simon's (1955) model proved useful, and contributes to the ongoing research agenda on the evolution of networks.

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 non-linear, but it is a Bernoulli equation so it can be linearized and integrated. We could also note that it is an exact differential equation and apply relevant techniques. The simplest is probably to separate variables to obtain

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

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

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

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

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

Note that equation A.2 is a logistic curve, that is, diffusion is S-shaped (from equation A.1, $\frac{d^2 k_j(t)}{dk^2}$ changes sign at $k_j = n/2$). The continuous distribution is computed thus (using equation A.2)

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

Since the t_j's are uniformly distributed[11] their probability mass function is $prob(t_j = Y) = 1/t$ for Y from 1 to t, so $prob(t_j \leq Y) = \sum_1^Y \frac{1}{t} = \frac{Y}{t}$. This leads to

$$p(k_j \leq k) = 1 - \left(\frac{k_j - k_j n}{k_j - n}\right)^{\hat{b}\mu}$$

Applying $p(k) = \frac{dp(k_j \leq k)}{dk}$ 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. Rewrite (A.3) as

$$p(K = k) \propto (n-k)^{-1+\hat{b}\mu}\ k^{-1-\hat{b}\mu}$$

Define the random variable $X = K/n$

$$p(X = x) = p(K = nx) \propto (n - nx)^{-1+\hat{b}\mu}\ (nx)^{-1-\hat{b}\mu}$$

The factors n are now absorbed into the constant

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

which is the definition of 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.

B Distribution of agents' number of ideas known

Below it is shown that the the number of ideas known by a r.c. agent has a binomial distribution. 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:

Proposition B.1. *The distribution of agents' number of ideas known is the binomial distribution*

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

[11] Contrary to one-mode scale free network models, this is not exactly true, since there is not one new paper 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 $[1, t]$.

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

Written explicitly the fixed point equation 14 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 $r_1 r_2 = -2\hat{b}\mu$. The series itself is thus $1 + 4\frac{1-2}{r_1 r_2} = \frac{\hat{b}\mu+2}{\hat{b}\mu}$. The term $p(1)$ is $(p_1|n = 2) = \frac{2\hat{b}\mu}{2\hat{b}\mu+1}$. Substituting into (C.1),

$$1 - \mu = \frac{b(2\hat{b}\mu)(\hat{b}\mu - 2)}{2(2\hat{b}\mu + 1)a} = \frac{b(\hat{b}\mu + 2)}{2\hat{b}\mu + 1}$$

that is

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

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

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

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 $P(j \in N_{i'} \backslash N_i) = \frac{2}{2n}$, so that$|N_{i'} \backslash N_i| = \frac{w}{n}$. The attachment kernel is then $A_t(k_j) = \frac{1-b}{bt}$, 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 $k \leq n - 1$, 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 $P(j \in N_{i'} \backslash N_i)$ 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

Beguerisse-Diaz, 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', *Proceedings of the National Academy of Sciences* **101**(Suppl 1), 5266.

Cohen, W. & Levinthal, D. (1989), 'Innovation and learning: the two faces of R & D', *The Economic Journal* **99**(397), 569–596.

Cowan, R. & Jonard, N. (2004), 'Network structure and the diffusion of knowledge', *Journal of Economic Dynamics and Control* **28**(8), 1557–1575.

Cowan, R. & Jonard, N. (2009), 'Knowledge portfolios and the organization of innovation networks', *The Academy of Management Review* **34**(2), 320–342.

Evans, T. & Plato, A. (2008), 'Network rewiring models', *Networks and Heterogenous Media* **3**(2), 221–238.

Geroski, P. A. (2000), 'Models of technology diffusion', *Research Policy* **29**(4), 603–625.

Gosper, R. W. (1978), 'Decision procedure for indefinite hypergeometric summation', *Proceedings of the National Academy of Sciences* **75**(1), 40–42.

Gutirrez Jimez, R. & Rodrguez Avi, J. (1997), 'Family of pearson discrete distributions generated by the univariate hypergeometric function 3f2', *Applied Stochastic Models and Data Analysis* **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?', *The American Economic Review* **97**(3), 890–915.

Johnson, N. L., Kemp, A. W. & Kotz, S. (2005), *Univariate discrete distributions*, Vol. 444, Wiley-Interscience.

Jovanovic, B. & Rob, R. (1989), 'The growth and diffusion of knowledge', *The Review of Economic Studies* **56**(4), 569–582.

König, M., Lorenz, J. & Zilibotti, F. (2012), 'Innovation vs imitation and the evolution of productivity distributions', *CEPR Discussion Paper No. DP8843* .

Krapivsky, P., Redner, S. & Leyvraz, F. (2000), 'Connectivity of growing random networks', *Physical Review Letters* **85**(21), 4629–4632.

Lafond, F. (2012), 'Learning and the structure of citation networks', *UNU-MERIT working paper* **2012-071**.

Liu, C., Yeung, C. & Zhang, Z. (2011), 'Self-organization in social tagging systems', *Physical Review E* **83**(6), 066104.

Lucas, R. E. J. (1988), 'On the mechanics of economic development', *Journal of Monetary Economics* **22**(1), 3–42.

Lucas, R. E. & Moll, B. (2013), 'Knowledge growth and the allocation of time', *Draft* .

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', *Journal of Symbolic Computation* **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', *EPL (Europhysics Letters)* **79**, 28001.

Petkovšek, M., Wilf, H. & Zeilberger, D. (1996), *A=B*, Wellesley, MA, http://www.math.rutgers.edu/ zeilberg/AeqB.pdf.

Price, D. (1976), 'A general theory of bibliometric and other cumulative advantage processes', *Journal of the American Society for Information Science* **27**(5), 292–306.

Ramasco, J., Dorogovtsev, S. & Pastor-Satorras, R. (2004), 'Self-organization of collaboration networks', *Physical Review E* **70**(3), 036106.

Romer, P. M. (1990), 'Endogenous technological change', *Journal of Political Economy* pp. 71–102.

Roth, C. & Cointet, J. (2010), 'Social and semantic coevolution in knowledge networks', *Social Networks* **32**(1), 16–29.

Schultz, T. W. (1961), 'Investment in human capital', *The American Economic Review* **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.

Sen, A. (2001), *Development as freedom*, Oxford Paperbacks.

Simon, H. (1955), 'On a class of skew distribution functions', *Biometrika* **42**(3/4), 425–440.

Slater, L. J. (1966), *Generalized hypergeometric functions*, University Press Cambridge.

Solow, R. M. (1957), 'Technical change and the aggregate production function', *The Review of Economics and Statistics* **39**(3), 312–320.

Vázquez, A. (2003), 'Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations', *Physical Review E* **67**(5), 056104.

Young, H. P. (2009), 'Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning', *The American Economic Review* **99**(5), 1899–1924.

Zeng, A., Yeung, C., Shang, M. & Zhang, Y. (2012), 'The reinforcing influence of recommendations on global diversification', *EPL (Europhysics Letters)* **97**(1), 18005.

The UNU-MERIT WORKING Paper Series

2013-01 *Effects of innovation on employment in Latin America* by Gustavo Crespi and Ezequiel Tacsir
2013-02 *Revisiting the porter hypothesis: An empirical analysis of green innovation for the Netherlands* George van Leeuwen and Pierre Mohnen
2013-03 *Impact of external knowledge acquisition strategies on innovation - A comparative study based on Dutch and Swiss panel data* by Spyros Arvanitis, Boris Lokshin, Pierre Mohnen and Martin Wörter
2013-04 *Interactive knowledge exchanges under complex social relations: A simulation modelRobin* by Cowan and Anant Kamath
2013-05 *Innovation systems framework: still useful in the new global context?* by Michiko Iizuka
2013-06 *The importance of intrinsic and extrinsic motivation for measuring IQ* by Lex Borghans, Huub Meijers and Bas ter Weel
2013-07 *Firms' innovation capability-building paths and the nature of changes in learning mechanisms: Multiple case-study evidence from an emerging economy* by Paulo N. Figueiredo , Marcela Cohen and Saulo Gomes
2013-08 *A set of time series data labour market stocks and flows for the Netherlands 1980 to 2010* by Manuel Müllers, Joan Muysken and Erik de Regt
2013-09 *Designing an optimal 'tech fix' path to global climate stability: R&D in a multi-phase climate policy framework* by Adriaan van Zon and Paul A. David
2013-10 *Complementarity between internal knowledge creation and external knowledge sourcing in developing countries* by Jun Hou and Pierre Mohnen
2013-11 *Summarizing large spatial datasets: Spatial principal components and spatial canonical correlation* by Samyukta Bhupathiraju, Bart Verspagen and Thomas Ziesemer
2013-12 *Regional systems of innovation in the Arab region* by Samia Satti Osman Mohamed Nour
2013-13 *Development and social justice: Education, training and health in Sudan* by Samia Satti Osman Mohamed Nour
2013-14 *The economic importance and impacts of intellectual property rights (IPRs) in Sudan* by Samia Satti Osman Mohamed Nour
2013-15 *Overview of knowledge economy in the Arab region* by Samia Satti Osman Mohamed Nour
2013-16 *The importance (impacts) of knowledge at the macro-micro levels in the Arab Gulf countries* by Samia Satti Osman Mohamed Nour
2013-17 *Political determinants and impact analysis of using a cable system as a complement to an urban transport system* by Diego Escobar-García, Francisco García-Orozco and Carlos Cadena-Gaitán
2013-18 *Women entrepreneurs in the informal economy: Is formalization the only solution for business sustainability?* By Shyama V. Ramani, Ajay Thutupalli, Tamas Medovarszki, Sutapa Chattopadhyay, Veena Ravichandran
2013-19 *Heterogeneity in innovation strategies, evolving consumer preferences and market structure: An evolutionary multi-agent based modelling approach* by Salih Çevikarslan

2013-20 *Optimal patent length and patent breadth in an R&D driven market with evolving consumer preferences: An evolutionary multi-agent based modelling approach* by Salih Çevikarslan
2013-21 *Innovation and productivity: An update* by Pierre Mohnen and Bronwyn H. Hall
2013-22 *Fathers' use of parental leave. What do we know?* by Nevena Zhelyazkova
2013-23 *Eliciting Illegal migration rates through list randomization* by David McKenzie and Melissa Siegel
2013-24 *How do ICT firms in Turkey manage innovation? Diversity in expertise versus diversity in markets* by Semih Akçomak, Erdal Akdeve and Derya Fındık
2013-25 *Dynamic models of R&D, innovation and productivity: Panel data evidence for Dutch and French manufacturing* by Wladimir Raymond, Jacques Mairesse, Pierre Mohnen and Franz Palm
2013-26 *Centre-based versus home-based childcare* by Robert Bauchmüller
2013-27 *Microeconometric evidence of financing frictions and innovative activity* by Amaresh K Tiwari, Pierre Mohnen, Franz C Palm and Sybrand Schim van der Loeff
2013-28 *Innovation for economic performance: The case of Latin American firms* by Elena Arias Ortiz, Gustavo Crespi, Ezequiel Tacsir, Fernando Vargas and Pluvia Zuñiga
2013-29 *Is money all? Financing versus knowledge and demand constraints to innovation* Gabriele Pellegrino and Maria Savona
2013-30 *Child deprivation in Ontario - A (less than perfect) comparison with Europe* by Geranda Notten
2013-31 *Measuring performance: does the assessment depend on the poverty proxy?* by Geranda Notten
2013-32 *How big is the impact of infrastructure on trade? Evidence from meta-analysis* by Mehmet Güney Celbis, Peter Nijkamp and Jacques Poot
2013-33 *Using a 'Systems' Perspective to Explain the Limits of 'New' Multinational Enterprises: the role of 'members-only' location advantages* by Rajneesh Narula
2013-34 *Foreign direct investment as a driver of industrial development: why is there so little evidence?* by Rajneesh Narula
2013-35 *The end of the multifibre arrangement (MFA) and the heterogeneous performance of quota-constrained countries* by Mulu Gebreeyesus
2013-36 *Techological capability building in MNE-related social businesses of less developed countries: The experience of Grameen-Danone Foods in Bangladesh* by Jahan A. Peerally and Paulo N. Figueiredo
2013-37 *The links between economic integration and remittances behaviour of migrants in the Netherlands* by Özge Bilgili
2013-38 *The influence of vulnerability on migration intentions in Afghanistan* by Craig Loschmann and Melissa Siegel
2013-39 *How unemployment insurance savings accounts affect employment duration: Evidence from Chile* by Paula Nagler
2013-40 *Self-organization of knowledge economies* by François Lafond