IL NUOVO CIMENTO **39 C** (2016) 280
DOI 10.1393/ncc/i2016-16280-3

COMMUNICATIONS: SIF Congress 2015

The dynamics of innovation through the expansion in the adjacent possible

F. TRIA(1)(2)

(1) *Dipartimento di Fisica, Sapienza Università di Roma - Roma, Italy*
(2) *ISI Foundation - Turin, Italy*

received 8 February 2016

Summary. — The experience of something new is part of our daily life. At different scales, innovation is also a crucial feature of many biological, technological and social systems. Recently, large databases witnessing human activities allowed the observation that novelties —such as the individual process of listening a song for the first time— and innovation processes —such as the fixation of new genes in a population of bacteria— share striking statistical regularities. We here indicate the expansion into the adjacent possible as a very general and powerful mechanism able to explain such regularities. Further, we will identify statistical signatures of the presence of the *expansion into the adjacent possible* in the analyzed datasets, and we will show that our modeling scheme is able to predict remarkably well these observations.

1. – Introduction

Innovation is a fundamental factor in the evolution of biological systems, human society and technology [1-3]. From this perspective, a deep understanding of the underlying mechanisms through which innovations emerge, diffuse, compete and stabilize is key in many different fields. Though a lot has been investigated about the way innovations can possibly emerge in specific sectors [4-7], the general picture remains poorly understood theoretically and undocumented empirically. The availability of extensive longitudinal records of human activity online [8] allows now to make reliable measures to highlight for common signatures of different systems where innovation occurs. This will set a common ground to test different intuitions and modeling schemes. An important observation is that we can analyse in the same framework both true innovations and simple novelties. They share in fact an important feature: they can be viewed as first-time occurrences of something at the individual or collective level, respectively. More precisely, an innovation is something created for the first time, something new to the world —*e.g.* a new gene, a new technology, a new word. A novelty, by contrast, is merely anything that is new

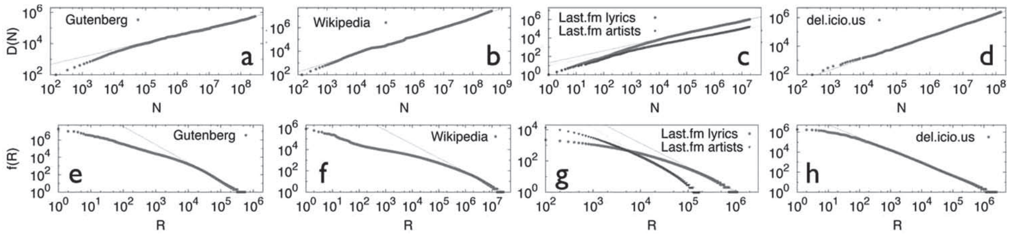

Fig. 1. – Heaps' law (top) and Zipf's law (bottom) in real datasets. *Gutenberg* `http://www.gutenberg.org/` (a,e): corpus of about 4600 English texts dealing with diverse subjects and including both prose and poetry, available at the Gutenberg Project ebook collection. *Wikipedia* (`http://www.wikipedia.org/`): The Wikipedia database we collected [17] dates back to March 7th, 2012 and contains a copy of all pages with all their edits in plain text. (b,f). *Last.fm* (`http://last.fm`) (c,g): is a music website equipped with a music recommender system. The data set we used [18,19] contains the whole listening habits of 1000 users until May, 5th 2009. `Del.icio.us` (`http://delicious.com/`) (d,h): is an online social annotation platform of bookmarking where users associate keywords (tags) to web resources (URLs) in a post. The dataset covers almost 3 years of user activity, from early 2004 up to November 2006. Straight lines in the Heaps' law plots show functions of the form $f(x) = ax^{\beta}$, with the exponent β equal respectively to $\beta = 0.45$ (Gutenberg), $\beta = 0.77$ (Wikipedia), $\beta = 0.68$ (Last.fm lyrics), $\beta = 0.56$ (Last.fm artist), and $\beta = 0.78$ (del.icio.us). Straight lines in the Zipf law plots show functions of the form $f(x) = ax^{-\alpha}$, where the exponent α is equal to β^{-1} for the different β's considered above. Note that the frequency-rank plots in real data deviate from a pure power-law behavior and the correspondence between the β and α exponents is valid only asymptotically.

to you (or to someone else) —*e.g.* an expression you start to use at some point, a song you listen for the first time, a technology you adopt. Seen in this light, innovations are novelties to everyone.

In order to build a bridge between the study of innovations in biological, technological and social systems, and the ubiquitous but often overlooked novelties we all experience every day, we here considered four data sets capturing various facets of innovation and novelties in social and technological systems. Each dataset consists of a sequence of elements ordered in time. (1) *Texts*: Here the elements are words. A novelty in this setting is defined to occur whenever a word appears for the first time in the text; (2) *Online music catalogues*: The elements are songs. A novelty occurs whenever a user listens either to a song or to an artist that she has not listened to before; (3) *Wikipedia*: The elements are individual wikipages. An innovation corresponds to the first edit action of a given wikipage by a given contributor (the edit can be the first ever, or other contributors may have edited the page previously but that particular contributor has not); (4) *Social annotation systems*: In the so-called tagging sites, the elements are tags (descriptive words assigned to photographs, files, bookmarks, or other pieces of information). A novelty/innovation corresponds either to the introduction of a brand new tag (a true innovation), or to its adoption (simple novelty) by a given user. We will show that all the considered datasets share common statistical features, regardless of the fact that they are dealing with true innovations or novelties.

The first statistical signatures we can consider are related to the rate at which novelties happen and to the frequency distribution of the elements in the observed sequence of events. In particular, the rate at which novelties occur can be quantified by focusing on the growth of the number $D(N)$ of distinct elements (words, songs, wikipages, tags) in a temporally ordered sequence of data of length N. Figure 1 (top) shows a sublinear power-law growth of $D(N)$ in all four data sets, each with its own exponent $\beta < 1$. This sublinear growth is the signature of Heaps' law [9]. It implies that the rate at which novelties occur decreases over time as $t^{\beta-1}$. The Zipf law [10] refer to the frequency-rank distribution of the different elements inside each sequence of data. In all cases (fig. 1, bottom) the tail of the frequency-rank plot follows an approximate power law, and its exponent α is compatible with the measured exponent β of Heaps' law for the same data set, via the well-known relation $\beta = 1/\alpha$ [11, 12]. We note that this relation is derived from a random-sampling argument: if one assumes a strict power-law behaviour of the frequency-rank distribution $f(R) \sim R^{-\alpha}$ and constructs a sequence by randomly sampling from this Zipf's distribution $f(R)$, one recovers Heaps' law with the functional form $D(t) \sim t^{\beta}$ with $\beta = 1/\alpha$. But the assumption of random sampling is strong and sometimes unrealistic. If one relaxes the hypothesis of random sampling from a power-law distribution, the relationship between Zipf's and Heaps' law becomes far from trivial. In this respect, it is important to observe that the frequency-rank plots in fig. 1 feature a variety of system-specific behaviors and the relation $\beta = 1/\alpha$ between the exponent β of Heaps' law and the exponent α of Zipf's law holds only in the tail of the Zipf's plot.

Despite the fact that Zipf's and Heaps' laws have been observed in many systems for a long time, and despite many attempts of accounting for this universality, (see for instance [11, 12]), a general mechanism able to independently explain both observations without any *a priori* hypotheses on their functional form or their relationship was yet to be elucidated. We recently introduced [13] an original mathematical model based on Pólya urns [14, 15] and on the concept of expansion into the adjacent possible [16], that predicts, in a very general framework, both the Zipf's and the Heaps laws, and their asymptotic relationship.

2. – Urn model with triggering

In a seminal paper [14] Pólya discussed a very simple but powerful modeling scheme to describe phenomena where the present state features a strong dependence on the past history. In the simplest version of the model, balls of two different colors are placed in an urn. A ball is withdrawn at random, inspected, and placed back in the urn along with a certain number of new balls of the same color, thereby increasing that color's likelihood of being drawn again in later rounds. This simple model, named the contagion scheme, captures an essential feature of many real phenomena, namely the global influence of past events on the evolution of the present situation. In 1984, Fred M. Hoppe [20] introduced for the first time innovation in the framework of Pólya urn models. The motivation of Hoppe's work was to derive in a simple and intuitive way the Ewens sampling formula [21], that describes the allelic partition at equilibrium of a sub-population evolved according to a discrete-time Wright-Fisher process [22,23] with constant mutation rate per gene. In order to mimic the mutation process, the model allows for brand-new colors to enter in the urn during the evolutionary process. This possibility is introduced through a special ball, the "mutator". In particular, the process starts with only the mutator in the urn. A ball is withdrawn at random, and, if the ball is the mutator, it is placed back in the urn along with a ball of a brand-new color, thus increasing the number of different colors present in the urn. Otherwise, the selected ball is placed back in the urn along with another ball of the same color. It is possible to show [20, 24, 25] that the Hoppe model predicts a logarithmic increase of new colors in the urn, resulting in an innovation rate far slower than the one observed in many real systems, as shown in the previous section. Hoppe's urn scheme is non-cooperative in the sense that one novelty does nothing to facilitate another. In other words, while in the Hoppe model a mechanism that allows the expansion of the space of possibilities is already present, this mechanism is completely independent on the actual realization of a novelty.

Here we propose a modeling scheme based on urn models and allowing for innovation, that incorporates the notion that by opening up new possibilities, innovations pave the way to other innovations. This idea is related to the Kauffman theoretical concept of the *adjacent possible* [26,16], the space of possibilities reachable from the present state, which he originally discussed in his investigations of molecular and biological evolution, and which has also been applied to the study of innovation and technological evolution [3,27].

2˙1. *Definition of the urn model with triggering.* – Consider an urn $\mathcal{U}$, representing the space of the possible, initially containing N_0 distinct elements, represented by balls of different colors. By randomly extracting elements from the urn, we construct a sequence $\mathcal{S}$ representing, among all possible histories, the path actually realized. Both the urn and the sequence enlarge during the process, as follows (refer to fig. 2, left). At each time step t, an element s_t is drawn at random from the urn, added to the sequence, and put back in the urn along with ρ additional copies of it (fig. 2 (left), A); leave as is the chosen element s_t is new (*i.e.*, it appears for the *first time* in the sequence $\mathcal{S}$), $\nu+1$ brand new distinct elements are also added to the urn (fig. 2 (left), B). These new elements represent the set of new possibilities opened up by the seed s_t. Hence $\nu+1$ is the size of the *new adjacent possible* available once an innovation occurs.

Our simple urn model with innovation accounts simultaneously for the emergence of both the Heaps and the Zipf laws (fig. 2, right). In particular, one can derive simple asymptotic formulas for the number $D(t)$ of distinct elements appearing in the sequence

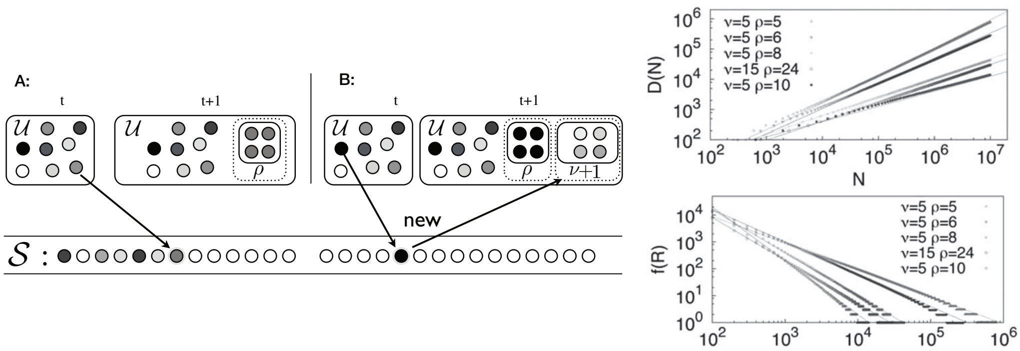

Fig. 2. – left: Urn model with triggering. A: An element that had previously been drawn from the urn is drawn again: the element is added to $\mathcal{S}$ and it is put back in the urn along with ρ additional copies of it. B: An element that never appeared in the sequence is drawn: the element is added to $\mathcal{S}$, put back in the urn along with ρ additional copies of it, and $\nu+1$ brand-new and distinct balls are also added to the urn. right: Heaps law (top) and Zipf's law (bottom) in the urn model with triggering. Straight lines in the Heaps law plots show functions of the form $f(x) = ax^{\gamma}$ with the exponent $\gamma = \nu/\rho$ as predicted by the analytic results and confirmed in the numerical simulations. Straight lines in the Zipf law plots show functions of the form $f(x) = ax^{-\alpha}$, with $\alpha = \gamma^{-1} = \rho/\nu$.

as a function of the sequence length t (Heaps' law), and for the asymptotic power-law behavior of the frequency-rank distribution (Zipf's law) in terms of the model parameters ρ and ν (refer to [13] or [25] for detailed calculations):

$$(1) \quad \rho > \nu \rightarrow D \sim (\rho-\nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}, \quad \rho < \nu \rightarrow D \sim \frac{\nu-\rho}{\nu+1} t, \quad \rho = \nu \rightarrow D \sim \frac{\nu}{\nu+1} \frac{t}{\log t},$$

and, for all the three cases, $f(R) = R^{-\frac{\rho}{\nu}}$. Thus, the exponents of the two power laws can be modulated by varying the relative strength of the innovation and the reinforcement mechanisms, reaching all the possible values.

We note that in the case $\rho < \nu$ we recover the results of the well-known Simon's model [28], originally proposed in the context of linguistics, that leads to a Zipf law $f(R) = R^{-(1-p)}$, where $p > 0$, and to a linear growth in time of the number of different words. The Zipf law with an exponent $\alpha \leq 1$ was also predicted in a recent modeling scheme [29], that proposed an alternative way to reinforcement to produce skewed frequency distributions [28]. The predictions for $\rho > \nu$, that correspond to what observed in real systems, are instead obtained in [13] for the first time, at the best of our knowledge, in the framework of a general microscopic dynamical model.

3. – Measuring the expansion into the adjacent possible

In the previous section we showed that the mechanism of the expansion into the adjacent possible allows to predict very general statistical signatures of systems where innovation events occur, namely the Zipf and the Heaps laws. We now want to push further our analysis to get more direct evidences of the presence of the proposed mechanism in real data. The urn model with triggering exploits the idea of the opening of new perspectives triggered by a novelty. We now want to investigate if a novelty also brings a bias towards the actual realization of the new possibilities it carries.

To investigate this process we need to introduce the notion of semantics, defined here as meaningful thematic relationships between elements. In fact, the new adjacent possible triggered by an innovation is not a bunch of unrelated elements, but consists on elements related to each others and to the innovation that has triggered them. For instance, in the case of music, one can imagine that when we first discover an artist or a composer, and we like it, it is very likely that we shall find out about her whole production unknown to us till then. This in turn can stimulate us to listen to other songs of the same artist and so on. In the general case, we can then consider semantic groups as groups of elements related by common properties. The actual definition of semantic groups depends on the data we are studying, and can be straightforward in some cases and ambiguous in others. To be more concrete, we will define the semantic groups in the datasets considered above: (i) in *Wikipedia* we can define a mother page of a given page (say s), as the page from where for the first time a link to s, was created both in the case when s was actually written (blue links), and in the case when s was only foreseen (red links). We can thus regard different pages as belonging to the same semantic group if they have the same mother page; (ii) in the *Online music catalogues (Last.fm)* different semantic groups for the listened songs can be identified with the corresponding song writers; (iii) for *texts* and *tags* there is not a straightforward definition of semantic groups, so we choose to consider each word/tag as bearing its own class.

With the above definitions, the triggering of novelties can be observed by looking at the non-trivial distribution of elements belonging to the same semantic groups. In

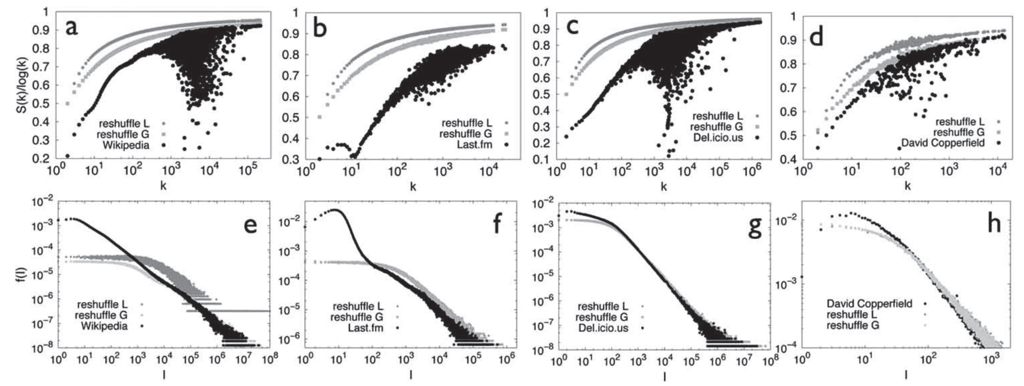

Fig. 3. – Normalized entropy (top) and distribution of time intervals (bottom) in the considered datasets. Results are displayed for Wikipedia (a,e), del.icio.us (b,f), Last.fm (c,g) datasets, and the David Copperfield book (d,h) (note that in this case it is only meaningful to consider correlations inside single books). In all the cases, results for the actual data are compared with the two null models, as described in the main text. The entropy is averaged for each k over the labels with the same number of occurrences.

particular, we observe whether elements belonging to the same semantic groups (in the following, we will refer to those elements as sharing the same label) occur in the sequence clustered together.

We introduce at this aim two specific observables.

(i) *The entropy S of a given label (say A).* It is defined by considering only the sub-sequence $\mathcal{S}_A$ starting from the first appearance of A. We divide $\mathcal{S}_A$ in a number of intervals k equal to the number of times the label A occurred in $\mathcal{S}_A$ (or equivalently in $\mathcal{S}$). Defining p_i as the frequency of the label A in each interval i, we can write the entropy

$$S_A(k) = -\sum_{i=1}^{k} p_i \log p_i. \tag{2}$$

We will then consider the normalised entropy $S_A(k)/S_A^{\max}(k) = S_A(k)/\ln k$.

(ii) *The distribution of time intervals $f(l)$* between two successive appearances of events with the same label.

In order to contrast the obtained results with a suitably defined random case, we consider two different ways of removing correlation in a sequence. Firstly, we just globally reshuffle the entire sequence $\mathcal{S}$ (we will refer to this procedure as global reshuffle). In this way, semantic correlations are disrupted but statistical correlations related to the non-stationarity of the model, responsible for instance of the Heaps and Zipf law, are still there. Secondly, for each label, we restrict to the sub-sequence $\mathcal{S}_A$ from the first appearance of that label, and then reshuffle it (we will call it local reshuffle). This latter procedure removes altogether any correlation between labels appearance. Figure 3 reports results for the normalised entropy (top) and the distribution of time intervals (bottom) as measured in the databases described above and in the globally and locally reshuffled sequence, showing a high-clustered behaviour in all the considered datasets.

4. – Urn model with semantic triggering

While the urn model with triggering predicts statistical correlations, it does not account for semantic correlations. In fact, observables measured on a sequence generated by the model show the same behaviour as when measured on a globally reshuffled sequence.

In order to account for the strong evidence of semantic correlations in real data, we need to generalise the previously introduced urn model with triggering by introducing the notion of semantics, and to introduce semantic correlations in the process of creating our sequence $\mathcal{S}$ of events.

4·1. *Model definition.* – By endowing balls (already carrying a color) with labels, we will consider the following generalization of the urn model with triggering. We start with an urn containing a number N_0 of distinct elements divided in $N_0/(\nu+1)$ groups, the elements in the same group sharing a common label. The urn and the sequence are then updated according to the following scheme (refer to fig. 4). The first element of the sequence $\mathcal{S}$ is randomly drawn from the urn, while at each time step t the weight of each element in the urn depends on the previous extraction in the following way: (i) a weight equal to 1 is given to each element with the same label as the previously extracted one (s_{t-1}), and to the element that triggered the entering in the urn of s_{t-1} (along with the other elements with the same label as s_{t-1}); (ii) a weight $\eta < 1$ is given to any other element in $\mathcal{U}$. The element s_t is drawn from $\mathcal{U}$ with a probability proportional to its weight. This introduces dynamical correlations into the process. After adding s_t to the sequence $\mathcal{S}$, it is put back in $\mathcal{U}$ along with ρ additional copy of it; iff the chosen element s_t is new (*i.e.*, it appears for the *first time* in the sequence $\mathcal{S}$) we also add to the urn $\nu+1$ brand new distinct elements, all with a common brand new label. Note that if $\eta = 1$ this model corresponds to the simple urn model with triggering introduced earlier.

In fig. 4 (left and center bottom respectively) we report results for the entropy and time-interval distribution as measured for sequences generated with the above process, for some specific values of the parameters ν, ρ and η. We can see that this generalized

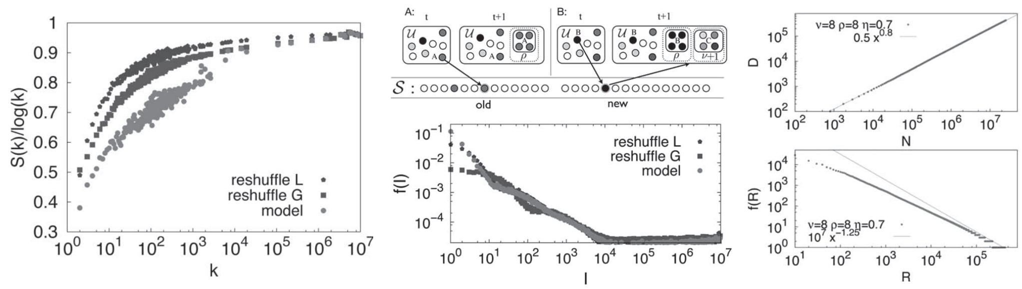

Fig. 4. – Urn model with semantic triggering. left: Normalized entropy: results are averaged over 10 realizations of the process, with parameters $\rho = 8$, $\nu = 10$, $\eta = 0.3$, $N_0 = \nu + 1$, and length $t = 10^7$. center: Top: model cartoon (refer to the main text for the explanation). Bottom: time-interval distribution, for the same data as for the entropy. right: Heaps (top) and Zipf's (bottom) laws for the parameters $\rho = 8$, $\nu = 8$, $\eta = 0.7$. The observed exponent for the Heaps law is within the theoretical bounds (refer to [13]) $\min(\frac{\nu\eta}{\rho}, 1) \leq \beta \leq \min(\frac{\nu}{\rho}, 1)$. The exponent of the tail of the frequency-rank distribution is compatible with the exponent of the Heaps law.

model with semantic triggering is able to reproduce the same qualitative behavior as found in the real systems analysed.

Figure 4 (right) also reports numerical results for the Heaps (top) and Zipf (bottom) laws. For this generalized model with semantic triggering, the relation between the exponent β of the Heaps law and the exponent $\alpha = 1/\beta$ of the Zipf law holds only asymptotically, *i.e.* for large times, with α measured on the tail of the frequency-rank distribution. We stress again that the existence of a pre-asymptotic regime for the Zipf law is observed also in real datasets. This suggests that taking into account correlations is crucial to explain the appearance of different regimes in the statistics of real datasets.

5. – Heaps' and Zipf's laws in the multicolors Pólya urn model

To fully appreciate the role of innovation as the underlying and almost universal force driving real systems, we now briefly discuss the Heaps and Zipf laws in the classical Pólya urn model with many colours, lacking innovation. Let us thus consider an urn initially containing N_0 balls, all of different colors (so that we have N_0 different colours). At each time step a ball is withdrawn at random, added to the sequence, and placed back in the urn along with ρ additional copies of it. Note that this process corresponds to the one depicted in fig. 2 (left), A.

The number of different colors $D(t)$ added to the sequence at time t follows the equation (when the continuous limit is taken)

$$(3) \qquad \frac{\mathrm{d}D}{\mathrm{d}t} = \frac{N_0 - D(t)}{N_0 + \rho t}, \quad D(0) = 0 \quad \Rightarrow \quad D(t) = N_0 \left[1 - \left(1 + \frac{\rho t}{N_0} \right)^{-\frac{1}{\rho}} \right],$$

where the solution is easily found by integration with separation of variables, and it is depicted in fig. 5 (left). Note that for $\rho t \ll N_0$, $D(t)$ follows a linear behaviour ($D(t) \simeq t$), while for large t it saturates at $D(t) \simeq N_0$, failing to predict a sublinear growth of new elements, as observed in many real systems.

Let us now compute the frequency probability distribution and the frequency-rank distribution for this model. The frequency probability distribution $p(f)$ can be obtained

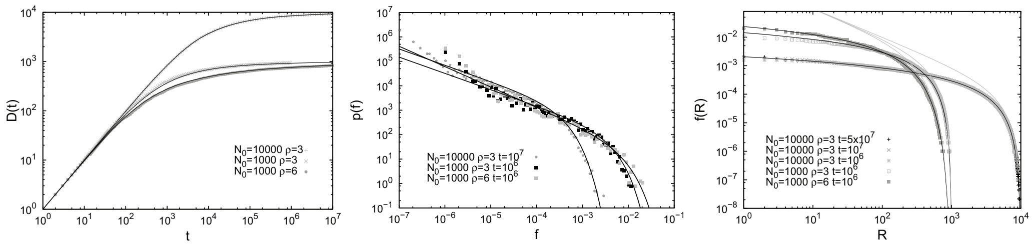

Fig. 5. – Results for the multicolors urn model without innovation. Results are reported both from simulations of the process (points) and from the analytical predictions (straight lines), for different values of the initial number of balls N_0 and of the reinforcement parameter ρ. left: Number of different colors $D(t)$ added in the sequence as a function of the total number t of extracted balls. The curves from analytical predictions of eq. (3) exactly overlap the simulated points. center: Frequency probability distribution. The straight line curves are predictions from eq. (4). right: Frequency-rank distribution. Simulations of the process are here reported along with both the predictions obtained by inverting eq. (5), and those from eq. (6), valid for $R \gg 1$.

by considering the frequency probability distribution in the long-time limit in the Pólya urn model with only two colors (refer for this to [14]), and initial conditions: N_0 balls in the urn, one of the color of which we are computing the limiting distribution, the other $N_0 - 1$ of the other color. We thus obtain (where ρ is again the reinforcement parameter)

$$p(f) = \frac{\Gamma\left(\frac{N_0}{\rho}\right)}{\Gamma\left(\frac{1}{\rho}\right)\Gamma\left(\frac{N_0-1}{\rho}\right)} f^{\frac{1}{\rho}-1}(1-f)^{\frac{N_0-1}{\rho}-1}. \tag{4}$$

Results are shown in fig. 5 (center). The frequency-rank curve (fig. 5, right) can now be obtained by inverting the relation

$$R \simeq N_0 \int_f^{+\infty} p(\tilde{f})\mathrm{d}\tilde{f} \simeq N_0 \int_f^{+\infty} C\tilde{f}^{\frac{1}{\rho}-1} e^{-\tilde{f}(\frac{N_0-1}{\rho}-1)}\mathrm{d}\tilde{f} = \frac{N_0}{\Gamma\left(\frac{1}{\rho}\right)}\Gamma\left(\frac{1}{\rho}, \frac{N_0-1-\rho}{\rho}f\right), \tag{5}$$

where in the last equality we used the approximation $f \ll 1$ (justified by the fast drop to zero of $p(f)$ for high values of f) and C is the normalization constant of the approximated distribution $C = \left(\frac{N_0-1}{\rho} - 1\right)^{1/\rho} \Gamma\left(\frac{1}{\rho}\right)^{-1}$.

We note that one could predict the behaviour of the tail of the frequency-rank distribution (for large values of the rank), by exploiting the relation between the Zipf and Heaps laws. In fact, in the long-time limit, we can consider the frequency of the element newly entered in the sequence as approximately inversely proportional to the sequence length. We can thus use the expression (3) for $D(t)$ and find the f that satisfies $f(D) = \frac{1}{t}$. We obtain

$$f(R) \simeq \frac{\rho}{N_0}\left[\left(1 - \frac{R}{N_0}\right)^{-r} - 1\right]^{-1}, \tag{6}$$

that well approximates the frequency-rank curve for $R \gg 1$ (refer to fig. 5, right).

6. – Conclusion

We presented a mathematical framework to describe systems where innovations (or novelties) occur, based on Pólya urns. We tested the predictions of our modeling scheme on large databases of human activities, quantifying the theoretical concept of the expansion into the adjacent possible. The considered datasets collect the efforts of many users, being therefore an evidence of a collective innovation dynamics. It is however possible to show [13] that the analysed statistical features are shared by the sequence of events created by single users. Generalization of our modeling scheme to include distinction between individual and collective dynamics could be of great interest to shed light on the dynamics of diffusion, fixation, success, turnover of popularity of novelties or innovations [30].

* * *

The author wishes to thank Vittorio Loreto, Vito D. P. Servedio and Steven Strogatz, with whom she realized the work presented here.

REFERENCES

[1] Jacob F., *Science*, **196** (1977) 1161.
[2] Kauffman S. A., *The Origins of Order: Self-Organization and Selection in Evolution* (Oxford University Press, New York) 1993.
[3] Johnson S., *Where Good Ideas Come From: The Natural History of Innovation* (Riverhead Hardcover, New York) 2010.
[4] Arthur W. B., *Econ. J.*, **99** (1989) 116.
[5] Dosi G., Ermoliev Y., (Yuriy) Y. and (Kaniovskyi) Kaniovski, *J. Math. Econ.*, **23** (1994) 1.
[6] Ziman J. M. (ed.), *Technological Innovation as an Evolutionary Process* (Cambridge University Press, Cambridge) 2000.
[7] Thurner S., Klimek P. and Hanel R., *New J. Phys.*, **12** (2010) 075029.
[8] Lazer D. *et al.*, *Science*, **323** (2009) 721.
[9] Heaps H. S., *Information Retrieval: Computational and Theoretical Aspects* (Academic Press, Inc., Orlando, FL, USA) 1978.
[10] Zipf G. K., *The Psychobiology of Language* (Houghton-Mifflin, New York, NY, USA) 1935.
[11] Serrano M. A., Flammini A. and Menczer F., *PLoS ONE*, **4** (2009) e5372.
[12] Lü L., Zhang Z. K. and Zhou T., *PLoS ONE*, **5** (2010) e14139.
[13] Tria F., Loreto V., Servedio V. D. P. and Strogatz S. H., *Sci. Rep.*, **4** (2014).
[14] Pólya G., *Ann. Henry Poincaré*, **1** (1930) 117.
[15] Johnson N. L. and Kotz S., *Urn Models and Their Application: An Approach to Modern Discrete Probability Theory* (John Wiley & Sons, Hoboken, NJ, USA) 1977.
[16] Kauffman S. A., *Investigations* (Oxford University Press, New York/Oxford) 2000.
[17] `http://dumps.wikipedia.org/enwiki/20120307/` (2012).
[18] Music recommendation datasets for research (2010). URL `http://goo.gl/lSZyib`.
[19] Celma O., *Music Recommendation and Discovery in the Long Tail* (Springer, Berlin) 2010.
[20] Hoppe F. M., *J. Math. Biol.*, **20** (1984) 91.
[21] Ewens W., *Theor. Popul. Biol.*, **3** (1972) 87.
[22] Fisher R. A., *The genetical theory of natural selection* (Clarendon Press, Oxford) 1930.
[23] Wright S., *Genetics*, **16** (1931) 97.
[24] Hoppe F. M., *J. Math. Biol.*, **25** (1987) 123.
[25] Loreto V., Servedio V. D. P., Strogatz S. H. and Tria F., in *"Universality and Creativity in Language"*, edited by Altmann E., Esposti M. Degli and Pachet F. *Lecture Notes in Morphogenesys* (Springer) 2015.
[26] Kauffman S. A., *Investigations: The Nature of Autonomous Agents and the Worlds They Mutually Create* SFI working papers (Santa Fe Institute) 1996.
[27] Wagner A. and Rosen W., *J. R. Soc. Interface*, **11** (2014) 20131190.
[28] Simon H. A., *Biometrika*, **42** (1955) 425.
[29] Corominas-Murtra B., Hanel R. and Thurner S., *Proc. Natl. Acad. Sci. USA*, **112** (2015) 5348.
[30] Monechi B., Ruiz-Serrano A., Tria F. and Loreto V., submitted (2015).