Crossing the horizon: exploring the adjacent possible in a cultural system

P. Gravino[1], B. Monechi[1], V. D. P. Servedio[2,3], F. Tria[3,1], V. Loreto[3,1,2]
[1]Institute for Scientific Interchange Foundation, Via Alassio 11/c, 10126, Turin, Italy
[2]Institute for Complex Systems (ISC-CNR), Via dei Taurini 19, 00185 Roma, Italy
[3]Sapienza University of Rome, Physics Dept., P.le Aldo Moro 2, 00185 Roma, Italy
pietro.gravino@gmail.com

Abstract

It is common opinion that many creative exploit are triggered by serendipity, fortuitous events leading to unintended consequences but this interpretation might simply be due to a poor understanding of the dynamics of creativity. Very little is known, in fact, about how innovations emerge and sample the space of potential novelties. This space is usually referred to as the *adjacent possible*, a concept originally introduced in the study of biological systems to indicate the set of possibilities that are one step away from what actually exists. In this paper we focus on the problem of portraying the adjacent possible space, and of analysing its dynamics, for a particular cultural system. We synthesised the graph emerging from the Internet Movies Database and looked at the static and dynamical properties of this network. We dealt with the subtle mechanism of the adjacent possible by measuring the expansion and the coverage of this elusive space during the global evolution of the system. We introduce the concept of adjacent possibilities at the level of single node to elucidate its nature by looking at the correlations with topological and user annotation metrics. We find that the exploration of the space of possibilities (potentially infinite by definition) shows a saturation size. Furthermore, single node analysis unveiled the importance of the adjacent possible as a useful probe for cultural impact.

The Invisible Horizon from the Shoulders of Giants

In a 1676 letter of Sir Isaac Newton can be found one of his most famous quotes: ”if I have seen further, it is by standing on the shoulders of giants”. With these words he meant to acknowledge and thank all the scholars that, with their efforts, made his work possible. The quote itself, actually, stems from at least four centuries before and was originally attributed to Bernard of Chartres. All cultural evolution processes strongly depend on the ability to stand on the shoulders of giants. Each new outcome of a cultural system is influenced by prior outcomes, just like in a biological system each offspring is the result of replications, recombinations and/or mutations of its ancestors DNA. The dynamics of evolution and innovation in cultural systems represents a very hot cross-disciplinary topic, which attracted several efforts from the scientific community in recent years (Mayer 1998; Elgammal and Saleh 2015; Tria et al. 2014; Jordanous, Allington, and Dueck 2015). In particular, the topic has been tackled form several angles: for example, by trying to understand and quantify the unexpectedness of commercial products (Grace and Maher 2014), by analysing the balance between originality and generativity in the creative cooperative production of online communities (Hill and Monroy-Hernández 2012) or by studying user linguistics behaviours and innovations on the web (Danescu-Niculescu-Mizil et al. 2013). These efforts have been made possible by the unprecedented availability of data tracking influences in the cultural activity typical of the Information Age we live in. Innovation phenomena do not just depend on the shoulders one is standing on. Innovators stand on the edge separating the previous knowledge from what still remains to be discovered. There is a wide horizon of innovations reachable from the verge of what is already known and, after Kauffman (1996), we name it as “adjacent possible”. By definition the adjacent possible gets continuously reshaped at every step forward in the unknown. We can describe cultural innovation processes like explorations in the hypothetical network of cultural entities linked by their influences (Wang, Song, and Barabási 2013; Spitz and Horvát 2014; Mauch et al. 2015). Though the way in which these influences are combined to produce novel outcomes is currently under the attention of scientists, very few attempts have been done, to the best of our knowledge, to analyse the way in which cultural network are explored so that the very notion of *adjacent possible* in cultural systems remains largely unexplored. Several question arise around this fascinating concept. How creative solutions do explore the adjacent possible frontiers? Do exploration patterns have long time lasting influence in the cultural network? Can this mechanism be improved to foster the insurgence of creative exploits? And, if so, how? In which way the creative exploration path covered in the past does influence future steps? Shedding some light on these questions could strongly improve our understanding of creativity and innovations both at an individual and at a societal level. This paper takes these lines of investigations by focusing on the the cultural system behind the cinematographic production. We adopted in particular a Web dataset of cinematographic production to reconstruct the network of influences among motion picture films. This network has been recently investigated (Wasserman, Zeng, and Amaral 2015; Spitz and Horvát 2014) with the aim to identify the most influential movies. Instead, here we focus on the notion of adjacent possible, both at the individual and collective level, with the aim of investigating its very definition and its structure as well as it gets explored by its community and reshaped over time. Though the adjacent possible remains a very elusive concept, a first portrait of its dynamics will emerge along with an interpretation of its meaning.

The Weaving of Influences in the History of Cinema

The Internet Movie Database (abbreviated IMDb, available at http://www.imdb.com) is an online database of information related to films, television programs and video games, including cast, production crew, fictional characters, biographies, plot summaries, trivia and reviews. The information comes from various sources. The IMDb team actively gathers information from studios and filmmakers though the the bulk of information is submitted by people in the industry and visitors. Sources of information include, though not limited to, on-screen credits, press kits, official bios, autobiographies, and interviews. Each movie web-page features metadata about awards, box office, releases date, plot keywords, ratings and connection between movies (spoofs, references, quotations, etc). In particular, the connections between pairs of movies are the crucial data we are interested in. To use them as a proxy for the movies influence, we downloaded the dataset of movies and connections, enriched with metadata about awards, ratings, etc., and then applied the following filtering procedure.

From the Raw Dataset to a Movie Influence Network

The IMDb dataset contains several millions of entities, many of which are not movies at all. The filtering procedure explained in the following, partly reproduces the work of (Wasserman, Zeng, and Amaral 2015). The platform contains information about TV shows, game show, news, videogames, music video and short movies and other formats. We reduced our analysis to "normal" movies, labeled in this way by the platform itself. Also, we considered movies with publication date in the period from 1909 to 2005. In this way we avoided the recentism of latest years productions, i.e., the tendency to over-annotate recent movies with respect to their historical importance. This over-annotation in connections would lead to a boosted high degree of some nodes that could bias the structure of the network.

Regarding the connections themselves, we adopted three of the eight types present in the dataset:

spoofs a fun reference to a title is made in a subsequent production;

features extracts from a title appear in another movie; e.g., a movie shows characters attending a cinema screening another movie, or the audio from a program is heard on a TV or a flashback sequence;

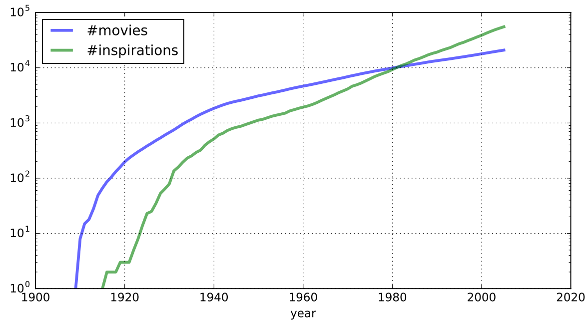

Figure 1: Growth in time of the number of movies and the number of inspiration links between them for the filtered graph.

references a title is referenced or a homage is paid to it in a subsequent movie; this includes recreations of movies scenes and off-screen references (e.g., the background music score)

The other five kinds of connections have been neglected because they are used much less frequently (ca. 10^3 times vs 10^5) and because they are mainly technical connections (e.g., re-edit or alternate language version). From the resulting set of movies and connections we constructed a direct graph (where the direction of links is chronological: influence moves from older movies to newer). Since time resolution is, in the worst cases, 1 year, we adopted this value as time resolution for every movie. We neglected all the interactions between movies of the same year. These interactions are usually unlikely in the dataset, and by doing so we get a tree structure, needed for our analysis. The graph resulting from this filtering has then been reduced to the largest weakly connected component. The final outcome is a graph, that we name the *inspiration graph*, with 20860 movies and 55219 links. The growth in time (by year) of the number of movies and of the number of connections is reported in Fig. 1. The links we are considering represent only the most explicit type of relation that can exists between two movies, without wanted to be exhaustive. Surely, influences are absolutely not limited to the ones reported in our dataset. The assumption we make is that our sample only captures the strongest relations among movies, somehow crucial for the development of a specific movie. In other words, we are assuming that a certain movie could not exist as it is without all the previously created ones with which it shares an inspiration link.

Properties of the inspiration graph

Before proceeding with the operative definition and the analysis of the adjacent possible, we report some basic analysis about the inspiration graph. Since we shall focus on the whole cinematographic system and its productions, it seems natural to consider the production itself, intended as the number of movies produced, as the intrinsic time of the dynamics. In this sense, the temporal unit of our system will be the creation of a movie, instead of the physical time. Fig. 2

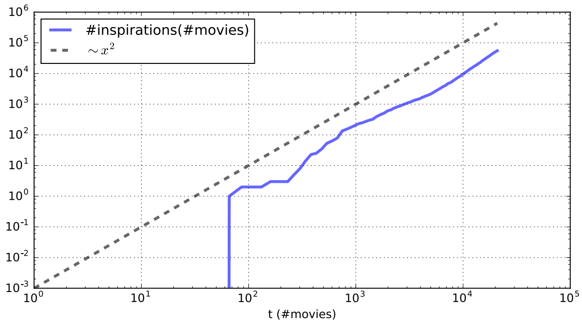

Figure 2: Growth of the number of inspiration links in the intrinsic time of the cultural system (the number of movies). A growing power law $\sim x^2$ is reported as a guide to the eye.

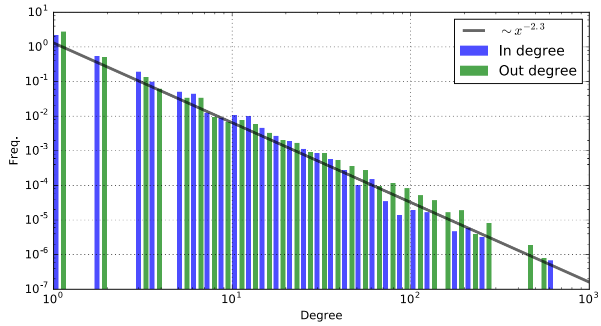

Figure 3: The histograms of the in- and out-degrees for the inspiration graph. The continuous line is the best fit with a power-law function.

reports again the growth of the network in this intrinsic time. The growth of the connections shows a steady power-law like growth (with exponent around 2) except for a few fluctuations, likely to represent the influence of historical, social and economical events (like World War II). An insight about the structure of the network is provided by the distribution of the in-degree (the number of influences received by a title) and of the out-degree (the number of influences coming from a title). Fig. 3 shows that unsurprisingly the distributions of these metrics can be described by power-law distributions. This kind of degree distribution is the signature of scale-free networks, which appear often in the analysis of human behaviour, in annotation process and in other well studied influences network (Spitz and Horvát 2014; Newman 2005; Wang, Song, and Barabási 2013). The distribution proved to be stable also against time resampling (e.g., by taking only a fraction of the story of the system), which means it is a stable feature consequent of the dynamic process we are analysing. The exponent of the power law has been estimated in ~ -2.3, with no significant difference between out- and in-degree distributions. We complete this preliminary analysis by looking at the distribution of time separations between related movies. The histogram of these distances is presented in Fig. 4, together with two

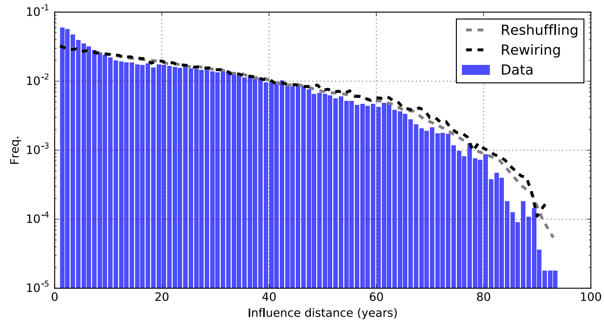

Figure 4: Histogram of the distances in years between related movies. As a reference, we reported the histograms of the distances of two null models: a random reshuffling of all edges and a rewiring preserving the degree distribution.

null models (a rewiring preserving degree distribution and a complete reshuffling of edges) (Albert and Barabási 2002; Wasserman, Zeng, and Amaral 2015). The comparison features a strong bias towards short temporal distances in real connections, which proved to be stable over time. This behaviour of the system highlights the natural tendency of movies to be influenced by those sharing the same cultural moment, like semantically correlated elements clustering in time (Tria et al. 2014).

The Adjacent Possible: Just One Step Away, in the Future

In this section we start by giving an operational definition of adjacent possible. Let us consider a generic graph of cultural productions linked by their influences with a dynamical process on it. At each time step, the graph can be divided in two parts: the known (or the actual) $K(t)$, i.e., the subset of nodes already explored, and the unknown (or the possible) $U(t)$, i.e., the subset of nodes still unseen. The exploration of this graph can only take place through influence links. We can thus define, at each time step, a subset of the unknown set containing all those nodes with all their influencers nodes belonging to the known set. This subset is defined as the "adjacent possible" at time t, $AP(t)$. Alternatively it can be defined as the set of unknown nodes that can be reached with the next step of exploration. An exemplification of the process is reported in Fig. 5. Since, by definition, the adjacent possible lies in the unknown part of the graph, we have no immediate access to it. Also, there is no guarantee that the future evolution of the system will reveal all the nodes belonging to the adjacent possible at any given moment. For sake of clarity let's consider an example. Suppose we are in 1950 and we look at the network of the whole production so far. In 1950 the adjacent possible of the nodes from 1930 ($AP(1930)$) will be represented by a given number of nodes (for instance the orange nodes in Fig. 5). If we now fast forward in time and land in the year 1980, we observe that the size of $AP(1930)$ will be larger, i.e., the number of orange nodes will have increased. This is a key point. The size of the observed adjacent possible depends on the point in time

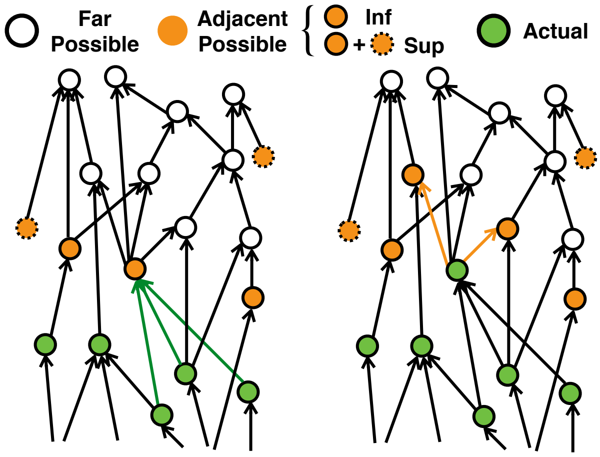

Figure 5: An exemplification of the exploration of the adjacent possible. The known nodes are in green (*actual*). Unknown and unaccessible nodes, i.e. with undiscovered inspirations, are in white (*far possible*). All the productions still unknown but with all the dependencies already discovered are in orange and represent the *adjacent possible*. Nodes with continuous contour have a non-zero in-degree, thus we know their main inspirations and their belonging to some specific adjacent possible, in its strict definition (*Inf*). Nodes with dashed contour do not have a in-degree and their inspiration are not known (they could be completely original or those inspirations could be simply not reported in the dataset or come from external media like books, news, etc). Thus, these nodes can be considered always in the loose definition of the adjacent possible (*Sup*), until they happen to be discovered. On the left, the graph before a production step. The new production is chosen among those in the adjacent possible. After the step, on the right, a new node is now known, and it has unlocked new nodes that are now part of the new adjacent possible.

from which we retrace the whole history. Presumably in 20 years time there will be new movies produced that will be adjacent adjacent to those of 1930. This means that, based on what we know and what we can measure, the adjacent possible could be an infinite set and it is only the finiteness of our sample that makes it finite. The best we can do is to measure the subset of adjacent possible observed at any given time. In practice what we can observe depends on two times: the time t at which we define the adjacent possible and the time t' ($t' > t$) from which we retrace the history. We can thus define the *observed adjacent possible* as:

$$\Gamma(t', t) = AP(t) \cap K(t') \tag{1}$$

where $K(t')$ is the set of known nodes at time t'. Though this set does not allow for a direct measure of $AP(t)$ it is very useful to provide us with valuable insights on how the exploration of $AP(t)$ takes place. Let us now apply this definition to our system. In our dataset we do not have the information about each intrinsic time step (i.e., each time a new movie comes out) since our time resolution is one year.

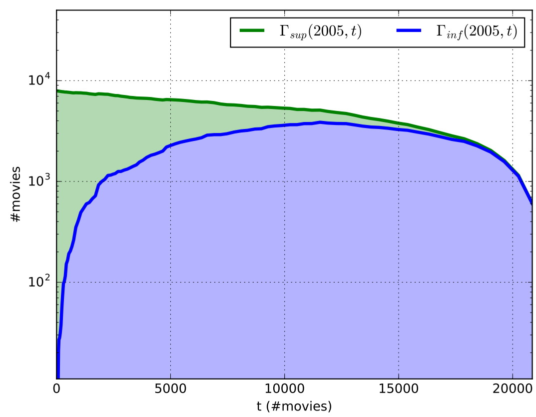

Figure 6: Measure of the superior ($\Gamma_{\text{sup}}(2005, t)$, in green) and inferior ($\Gamma_{\text{inf}}(2005, t)$, in blue) estimates for the observed adjacent possible of the inspiration network vs. the intrinsic time t of the system, i.e., the number of movies produced.

Still, we can define the state of knowledge of the network once a year, and consequently, we can estimate $\Gamma(t', t)$.

Before proceeding we should consider another element. In order for a node to be in the adjacent possible of other nodes, it must receive at least one influence, which means that the in-degree must be larger than 0. However, since the in-degree is distributed according to a power law, $k_{\text{in}} = 0$ is not only possible but is the most likely value. Actually, we cannot consider all these nodes as not having any influences at all. It is more likely that those influences have not been tracked yet or they come from sources external to our network (e.g., a book, a song, etc.). In order to overcome this problem, we define two metrics for the adjacent possible, depending on how we choose to treat nodes with $k_{\text{in}} = 0$. We can consider them as potentially uninfluenced, and then always in the adjacent possible until they happen to become part of the K set or we can simply neglect them. In one case, we are overestimating the size of the observed adjacent possible we can access, in the other case we are underestimating it. These ideas are explained in Fig. 5. We named these two metrics Γ_{sup} and Γ_{inf} and we measured both for each yearly state of knowledge of the network. Results are shown in Fig. 6. The measure gives us a general information about the typical size of Γ, which lies between 10^3 and 10^4 for the whole evolution. The measure in the final part looses reliability due to size effects, but still we can suppose that the size of Γ does not diverge with the size of the system. Let us now study directly the evolution of the coverage of the observed adjacent possible at a given time t. With our data, the best estimation that can be given of how the adjacent possible is going to be known during the exploration is to measure the evolution in time t of $\Gamma(2005, t') \cap K(t)$, i.e., the number of movies of the observed adjacent possible that are actually realized at each time t. In other words, with our data the best estimation for the adjacent possible of a given year t'

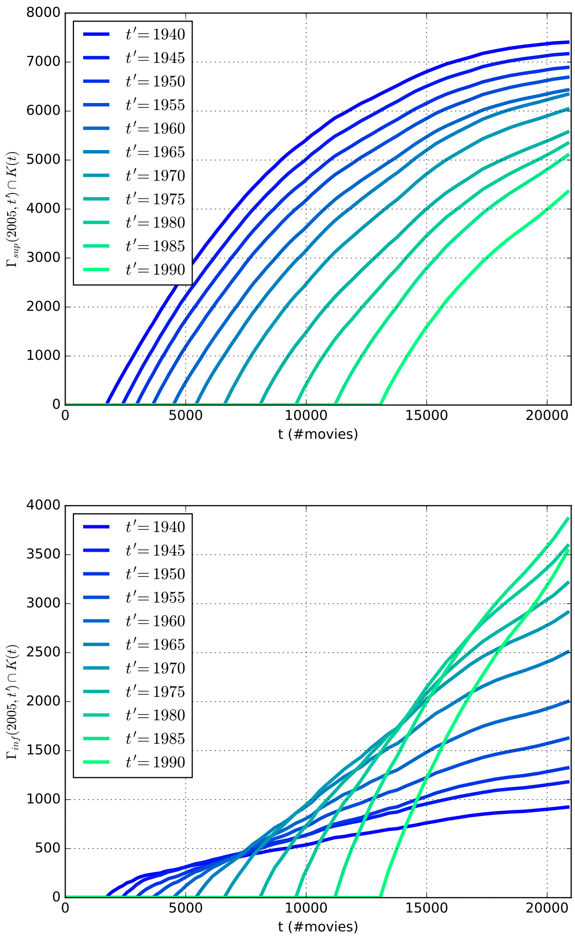

Figure 7: The evolution of the coverage of the observed adjacent possible Γ_{sup} (upper figure) and Γ_{inf} (lower figure). Different curves correspond to different values of t'.

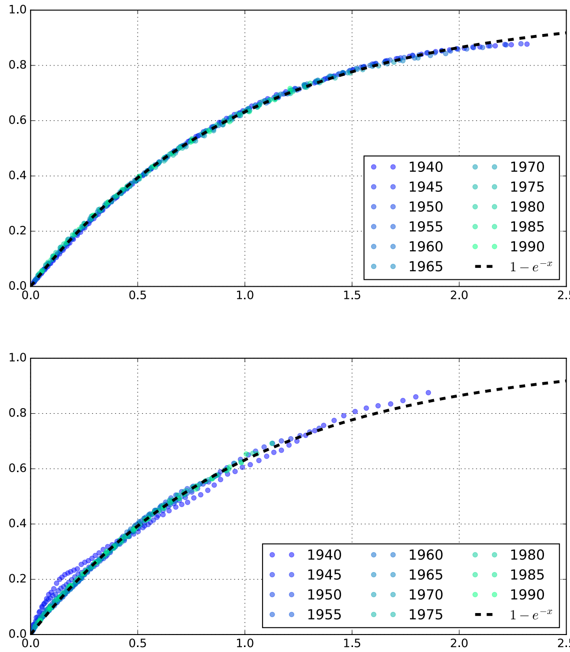

Figure 8: Collapse of the observed adjacent possible coverage curves of Fig. 7 when shifted and rescaled according the exponential fit parameters. Γ_{sup} curves are reported in the upper panel, while Γ_{inf} in the lower.

is the *observed adjacent possible* calculated using the whole timespan, i.e. updated to 2005. This set, according to Eq. 1, can be indicated with $\Gamma(2005, t')$. What we want to measure is how many movies of this set have been actualized in time t (where obviously $t > t'$). The results, for both metrics (Γ_{inf} and Γ_{sup}) of the observed adjacent possible, are reported in Fig. 7. Both measures, in particular Γ_{inf}, seem to show a tendency to saturation if a sufficient elapse of time has passed. To quantitatively account for this effect we fitted both types of curves with a function of the kind $y = a(1 - e^{-x/b})$, describing and exponential asymptotic relaxation towards a constant value defined by a. Fit results are not reported for the sake of brevity. Instead, we show in Fig. 8 how all the curves in Fig. 7 collapse when shifted and rescaled according to the transformations $x \to x/b$ and $y \to y/a$. We observe a convincing collapse for Γ_{sup} curves while Γ_{inf} curves feature some fluctuations around the master curve. To investigate such fluctuations, we fitted the curves of Γ_{sup} and Γ_{inf} for every time t'. Figure 9 shows the obtained fitting parameters a and b as a function of the intrinsic time (i.e., number of movies). It can be seen that for Γ_{inf} the largest fluctuations correspond to curves related to the first years of the dynamics. These measures are important because they tell us that even if the adjacent possible could be infinite, the *observed adjacent possible* of a given state of the inspirations network is covered in time in a way that suggests its boundedness. Indeed, and it was not obvious a priori, its discovered size seems to converge. Moreover, we have a quantitative account of time scales to reliably observe the convergence of the observed adjacent possible, or at least to estimate its size. The upper part of Fig. 9 shows the evolution with the growth of the system of the b parameter which is the time-scale of the exponential function fitting the coverage curves of Fig. 7. The parameter b can also be interpreted as the order of magnitude of the intrinsic time one should wait to have a reliable observation. Considering Γ_{sup} we see that the behaviour of both a and b as functions of the intrinsic time changes around a time $\approx 10^4$. Hence, considering this change as an effect of the finite size of the system, we are able to estimate correctly the size of Γ_{sup} for at least half of our dataset (i.e., every movie produced before $t \approx 10^4$). The evolution of b relative to Γ_{inf} tells us a different story. There is a clear peak (apparently limited only by the size of the dataset) representing a divergence of the timescale. Accordingly, since to accurately measure size we need data spanning more than a timescale, we can then conclude that size measures in the time frame of the di-

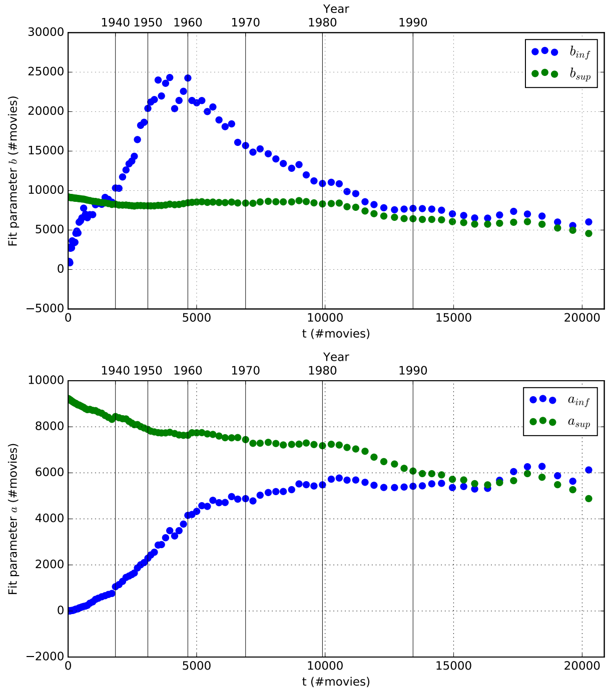

Figure 9: The evolution of exponential fitting parameters a and b. Each dot is a year for which we calculated the observed adjacent possible, measured the coverage curve (like in Fig. 7) and estimated the exponential fit parameters a and b. The curves represent the evolution of the coverage functions of the observed adjacent possible. In the upper figure, the sequence of values for the parameter b, representing the scale of time, both for Γ_{sup} (green) and for Γ_{inf} (blue); in the lower panel, the analogous for the parameter a, measured in number of movies.

vergence are less reliable. Comparing this with Fig. 1 and Fig. 2 we notice that such period is characterised by a strong change in the dynamics, which could have led the system to this instability. Looking at Fig. 9, we can see that this effect disappears for $t \sim 10^3$ (approximately half of the story covered by the dataset), where the curve lies smoothly in the same range of the b_{sup} curve. Looking instead at the size parameter a we observe, for Γ_{sup}, a decreasing curve with some discontinuity in the slope around 10^4. The asymptotic size of the observed adjacent possible seem not to be divergent with the size of the system and, thus, measurable (roughly estimate around $\sim 7 \cdot 10^3$). For the Γ_{inf} scale parameter we observe a different behaviour. A rapid growth until the peak of the unstable zone ($\sim 5 \cdot 10^3$) followed by a more or less stable plateau slightly under $\sim 6 \cdot 10^3$. In the last part of the evolution, the two parameters estimating the asymptotic size of the observed adjacent possible basically collapse, suggesting an high reliability of the measure.

Possible Meanings for the Adjacent Possible

The procedures implemented so far have, amongst the other purposes, the aim to prove the measurability of the observed

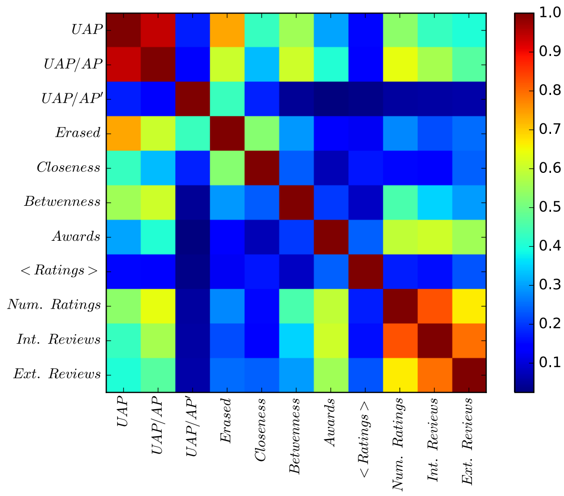

Figure 10: The matrix of the Pearson's coefficient between the adopted metrics.

adjacent possible and to give a quantitative insight about it. A qualitative understanding is also needed. To look more in detail at the possible meanings of the entity we defined at the global systemic scale, we can give also a microscopic definition of the adjacent possible of single nodes. In particular, if we consider Fig. 5, we can notice how the passage of a node from the adjacent possible to the known "unlocks" other nodes that, after the step, have become accessible and thus now belong to the adjacent possible. So, we can define for each node a metric depending on the unlocked adjacent possible (UAP in the following, referring to the observed adjacent possible). This metric, normalized in different ways, has to be compared with others, as described in the following. The comparison has been performed by calculating the Pearson's coefficient. The matrix of the results is reported in Fig. 10. The metrics evaluated are reported in the following, and are always relative to a generic node n.

UAP the number of nodes made available by the production of n; the number of nodes which were missing only n as reference to be in the adjacent possible.

UAP/AP$_{\text{sup}}$ the UAP metric normalised by the superior limit to the size of the adjacent possible observed for the year of n. AP_{sup} stands for $\Gamma_{\text{sup}}(2005, t(n))$.

UAP/AP$_{\text{inf}}$ the UAP metric normalised by the superior limit to the size of the adjacent possible observed for the year of n. AP_{inf} stands for $\Gamma_{\text{inf}}(2005, t(n))$.

Erased the number of nodes that would be unaccessible if the node n would never be discovered. This number has been estimated with the following algorithm.

1. we remove the node n;
2. we considered each node m amongst those influenced by n. We will assume that if n would not exist each node m would risk not to be discovered, depending on the importance of the influence between n and the specific m;

3. to estimate this importance, we consider all influences received by the specific m. We weight each influence received by m as the inverse of the in-degree of the node m. If the weight of the influence between n and m crosses a given threshold (30%), the influenced node m is removed and all its descendants will be considered for removal; i.e. if a node m has only two influences including n then the importance of the influence of n can be roughly estimated to be around $\sim 50\%$; thus, since this value exceedes the threshold, m will be removed and the movies influenced by m will have to be checked;
4. all nodes in the list of those to be considered for removal are analysed chronologically with the same rules, removing them if the sum of the weights of the deleted influences passes the threshold and adding their influenced nodes to the list, in case of removal.

This metric is an adapted version of the vitality metrics from (Brandes and Erlebach 2005).

Closeness Closeness centrality (Freeman 1978) is the reciprocal of the sum of the shortest path distances from n to all $N-1$ other nodes. Since the sum of distances depends on the number of nodes in the graph, closeness is normalised by the sum of minimum possible distances $N-1$.

$$C(n) = \frac{N-1}{\Sigma_m d(m,n)} \tag{2}$$

Betweenness The betweenness centrality (Brandes 2008; Brandes and Erlebach 2005) of a node n is the sum of the fraction of all-pairs shortest paths that pass through n:

$$c_B(n) = \sum_{m,m' \in V} \frac{\sigma(m,m'|n)}{\sigma(m,m')} \tag{3}$$

where V is the set of nodes, $\sigma(m,m')$ is the number of shortest (m,m')-paths, and $\sigma(m,m'|v)$ is the number of those paths passing through $n \neq m, m'$.

Awards The number of awards and nominations obtained by n as reported on the IMDb platform.

$\langle$*Ratings*$\rangle$ The average vote (from 1 to 10) for the movie n given by IMDb platform registered users.

Num. Ratings The number of votes for the movie n given by IMDb platform registered users.

Int. Ratings The number of reviews for the movie n submitted by IMDb platform registered users.

Ext. Ratings The number of links to external website reviews (usually from major print or online media organisations). Links are submitted on the IMDb platform by film reviewer or editor or a movie site team.

Let us now discuss the correlation matrix reported in Fig. 10. Our main matrix, UAP strongly correlates with UAP/AP_{sup} but very poorly with UAP/AP_{inf}. The latter shows a weak degree of correlation with the $Erased$ metric. This metric has been introduced to prove that UAP, despite its very local definition, can have long temporal range consequences. In fact, all the first three metrics show a strong correlation with it, meaning that the influence of a node with high UAP metric can reach high temporal distances, more or less directly. Closeness and betweenness centrality measures correlate fairly with the UAP and UAP/AP_{sup}. This is an insight of their value in the identification of nodes important from a topological point of view, connecting different communities or standing in the core of the network. All the other metric correlations proves the cultural value of nodes with high UAP and of UAP/AP_{sup} metric. It is worth to note the weakness of the correlation with $\langle$*Ratings*$\rangle$. This seems to suggest that the cultural value we are observing deals more with the interest gathered that with the appreciation (according to interest and appreciation information given by IMDb users).

Conclusions and perspective: the adjacent possible of this paper

In this paper we analysed the adjacent possible, the space in which creative efforts can move a step over the frontiers of what is known. We synthesised a network of influences between the entities of a cultural system. In particular we dealt with the cinematographic production system by leveraging the data extracted from the IMDb platform. With a suitable filtering procedure we sketched a graph of the most important influences and studied its structure and dynamical properties. In particular, we observed that despite the fact that the system showed an unstable growth rate, it resulted in a scale-free network of influences among movies. Moreover, these influences were found to be preferentially attached over short time distances (as inferred by comparing with null models). We then defined the observable projection of the adjacent possible according to the temporal resolution of one year. For each year, the observed adjacent possible was considered as the set of movies not yet produced whose inspirations lay all in the past. We had to define two kinds of observed adjacent possible in order to take into account nodes without annotated influences, the upper and lower bounds of the adjacent possible. We measured the adjacent possible for every year, within the dataset limits. Then we tracked how the adjacent possible of each year was covered by what was already known at previous times. This evolution led us to fit the coverage curves, and to estimate the typical time scale and the asymptotic limit for the size of the observed adjacent possible. Both numbers are, in the majority of cases, substantially smaller than the size and characteristic times of the whole network. This seems to suggest the existence of a saturation in the size of the observed adjacent possible at any given time that will be eventually explored. In other words, this result indicates that, even though the adjacent possible of a given state of the network is potentially unbounded, only a finite part of it is likely to be visited, and the size of this part can be estimated in a finite amount of time (e.g., with datasets of other cultural systems with a longer timespan or through computer simulations). This result is somehow surprising given our absolute, though natural, ignorance about the structure of the adjacent possible. Again, it is worth remarking that our conclusions apply to those parts of the network space one can observe, i.e., to the way in which that space was explored in history. In the last part of the paper we re-elaborated the definition of a suitable metrics for nodes, to be compared with other metrics, already known in literature or used by the IMDb dataset itself, related to the influence or the popularity of a movie. The metric we propose consists in the size of the unlocked adjacent possible (UAP). After a new node n is produced, the UAP is the number of nodes that were unreachable before and now are made available for production as a result of all known influences, including that of n itself. This metric, despite its local definition, was shown to be strongly correlated with a metric calculated on large temporal distances. Also, comparisons with standard topological metrics showed that high UAP values correspond to crucial nodes in the structure of the network. Finally, we also confirmed the cultural importance of the UAP as it correlates with the IMDb metrics, which are interesting for users. All these correlations confirmed the strategical importance of the adjacent possible concept even at the single node level. Thus, the study and understanding of its dynamic could be strategically fundamental to get a deeper comprehension of cultural system dynamics and evolution. The obvious problem for this is the time limit of the available statistics. This could be easily overcome by creating a model faithful enough to reproduce not only the statistical markers of the influence network but also the pattern of exploration of the adjacent possible. Given the peculiar characteristics of the network of influence this seems not to be an easy task, because the right balance between short time biases and preferential attachment (leading to a scale-free distribution) could be conflicting. Even when correctly balanced, there is no guarantee that the model would reproduce the correct adjacent possible exploration pattern. However, in case of success, such a model could confirm (or discard) our findings and could provide several answers about how creativity works and, maybe, can be improved at an individual and at a societal level. In fact, our metrics give us a new instrument to evaluate the value and the impact of creative productions. Also, this work can be considered as a first step toward a possible optimisation strategy for the exploration of the unknown. In fact, a deeper understanding of the *adjacent possible* exploration patterns could help to recreate opportune condition for a faster insurgence and spreading of creative solutions. We could understand if it is possible to efficiently drive innovation toward a given direction, and how, and this could completely transform, for example, our scientific research funding policies and our artistic or technologic evolution cycles. It is likely that a good *theory of the adjacent possible*, capable of such wonders, lies still far from *our actual adjacent possible*, but we hope our work could move the boundaries a bit toward that direction.

Acknowledgments

We acknowledge support from the KREYON project funded by the Templeton Foundation under contract n. 51663. VDPS acknowledges the EU FP7 Grant 611272 (project GROWTHCOM) and the CNR PNR Project "CRISIS Lab" for financial support.

References

Albert, R., and Barabási, A.-L. 2002. Statistical mechanics of complex networks. *Reviews of modern physics* 74(1):47.

Brandes, U., and Erlebach, T. 2005. *Network analysis: methodological foundations*, volume 3418. Springer Science & Business Media.

Brandes, U. 2008. On variants of shortest-path betweenness centrality and their generic computation. *Social Networks* 30(2):136–145.

Danescu-Niculescu-Mizil, C.; West, R.; Jurafsky, D.; Leskovec, J.; and Potts, C. 2013. No country for old members: User lifecycle and linguistic change in online communities. *WWW'13*.

Elgammal, A., and Saleh, B. 2015. Quantifying creativity in art networks. *arXiv preprint arXiv:1506.00711*.

Freeman, L. C. 1978. Centrality in social networks conceptual clarification. *Social networks* 1(3):215–239.

Grace, K., and Maher, M. L. 2014. What to expect when you're expecting: the role of unexpectedness in computationally evaluating creativity. In *Proceedings of the 4th International Conference on Computational Creativity, to appear*.

Hill, B. M., and Monroy-Hernández, A. 2012. The remixing dilemma: The trade-off between generativity and originality. *American Behavioral Scientist* 57(5).

Jordanous, A.; Allington, D.; and Dueck, B. 2015. Measuring cultural value using social network analysis: a case study on valuing electronic musicians. In *Proceedings of the Sixth International Conference on Computational Creativity June*, 110.

Kauffman, S. A. 1996. Investigations on the character of autonomous agents and the worlds they mutually create. In *Investigations*. Santa Fe Institute.

Mauch, M.; MacCallum, R. M.; Levy, M.; and Leroi, A. M. 2015. The evolution of popular music: Usa 1960–2010. *Royal Society open science* 2(5):150081.

Mayer, R. E. 1998. Fifty years of creativity research. In Sternberg, R. J., ed., *Handbook of Creativity*. Cambridge: Cambridge University Press. 449–460.

Newman, M. E. 2005. Power laws, pareto distributions and zipf's law. *Contemporary physics* 46(5):323–351.

Spitz, A., and Horvát, E.-Á. 2014. Measuring long-term impact based on network centrality: Unraveling cinematic citations. *PloS one* 9(10):e108857.

Tria, F.; Loreto, V.; Servedio, V. D. P.; and Strogatz, S. H. 2014. The dynamics of correlated novelties. *Scientific Reports* 4:5890.

Wang, D.; Song, C.; and Barabási, A.-L. 2013. Quantifying long-term scientific impact. *Science* 342(6154):127–132.

Wasserman, M.; Zeng, X. H. T.; and Amaral, L. A. N. 2015. Cross-evaluation of metrics to estimate the significance of creative works. *Proceedings of the National Academy of Sciences* 112(5):1281–1286.