

Interacting Discovery Processes on Complex Networks

Iacopo Iacopini^{1,2,3,*} Gabriele Di Bona^{1,4} Enrico Ubaldi⁵ Vittorio Loreto^{5,6,7} and Vito Latora^{1,3,8,7,†}

¹*School of Mathematical Sciences, Queen Mary University of London, London E1 4NS, United Kingdom*

²*Centre for Advanced Spatial Analysis, University College London, London W1T 4TJ, United Kingdom*

³*The Alan Turing Institute, The British Library, London NW1 2DB, United Kingdom*

⁴*Scuola Superiore di Catania, Università di Catania, Via Valdisavoia 9, 95123 Catania, Italy*

⁵*Sony Computer Science Laboratories, 6 Rue Amyot, 75005 Paris, France*

⁶*Sapienza University of Rome, Physics Department, Piazzale Aldo Moro 5, 00185 Rome, Italy*

⁷*Complexity Science Hub Vienna, A-1080 Vienna, Austria*

⁸*Dipartimento di Fisica ed Astronomia, Università di Catania and INFN, I-95123 Catania, Italy*



(Received 13 May 2020; revised 22 October 2020; accepted 5 November 2020; published 10 December 2020)

Innovation is the driving force of human progress. Recent urn models reproduce well the dynamics through which the discovery of a novelty may trigger further ones, in an expanding space of opportunities, but neglect the effects of social interactions. Here we focus on the mechanisms of collective exploration, and we propose a model in which many urns, representing different explorers, are coupled through the links of a social network and exploit opportunities coming from their contacts. We study different network structures showing, both analytically and numerically, that the pace of discovery of an explorer depends on its centrality in the social network. Our model sheds light on the role that social structures play in discovery processes.

DOI: 10.1103/PhysRevLett.125.248301

Discoveries are essential milestones for the progress of our societies [1–11]. Recently, different mathematical approaches have been proposed to model the dynamics of innovation [12–23]. Among these, of particular interest are those based on random processes with reinforcement [24–26], such as Pólya urns [27,28]. Urns have been extensively used to study and model a variety of systems and processes, from evolutionary economics, voting, and contagions [29–32] to language and folksonomies [33,34]. More recently, they have been employed to filter information [35] and grow social networks [23]. Interestingly, urns can also be used to model discovery processes, if opportunely combined with the concept of the *adjacent possible* (AP)—*the set of all those things which are one step away from what is already known* (Kauffman [36]). This formulation of the AP, which dates back to concepts previously introduced by Farmer, Langton, and others [37–39], has been translated into the urn model with triggering (UMT), a particular process in which the space expands together with the discovery dynamics, and the appearance of a novelty opens up the possibilities of further discoveries [4,40–43]. UMTs could successfully replicate the basic signatures of real-world discovery processes, such

as the famous Heaps’ and Zipf’s laws [44,45], often recurrent in complex systems [15,46–50], as well as Taylor’s law [51]. It turns out that the Heaps’ law, a sublinear growth of the number of distinct elements $D(t) \sim t^\beta$ with the number of elements t , well describes the pace at which scientists discover concepts or users collect new items [40,52,53], with higher values of β denoting a faster exploration of the AP. However, despite the fact that the existing models can capture essential underlying mechanisms behind the discovery of novelties, little emphasis is given to the collective dynamics of exploration and to the benefits that social interactions could bring. In fact, with the exception of Ref. [23], the modeled exploration dynamics refers to a single entity, representing, for example, the joint effort of researchers within a field [52]. Without taking into account the multiagent nature of the process, these models (i) do not capture the heterogeneity of the pace of the individual explorers and (ii) do not include the benefits brought by social interactions and collaborations. Indeed, empirical evidence of these mechanisms has been found in various contexts [54–56], such as music listening, politics, voting, and language [57–59].

In this Letter, we propose a model of interacting discovery processes where an explorer is associated to each of the nodes of a social network [60–62], and its dynamics is governed by an UMT. Hence, the local dynamics of each node accounts for the presence of an AP, more precisely, the *adjacent possible in the space of concepts*. The social network makes the exploration a

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article’s title, journal citation, and DOI.

collective one, since processes of neighboring urns are coupled. This coupling expands the notion of the AP by adding a social dimension, represented by the set of opportunities one is possibly exposed to through his or her social contacts. We call this the *adjacent possible in the social space*. Social networks have been extensively used as a substrate on top of which dynamical processes take place [63,64]. Notice, however, that our setting crucially differs from the typical approach in which the network mediates, for example, the diffusion of innovations or social contagions [65,66]. Here, the interactions among the many discovery processes reveals the twofold nature of the AP of each individual, highlighting the crucial role played by the social structure in determining the individual exploration dynamics.

Model.—Let us consider an unweighted directed graph $G(\mathcal{N}, \mathcal{E})$, where \mathcal{N} and \mathcal{E} are, respectively, a set of $N = |\mathcal{N}|$ nodes and a set of $E = |\mathcal{E}|$ links. Each node of the graph represents an individual or agent, while link (i, j) denotes the existence of a directed social relation from individual i to j (such that i can benefit from j). The graph is described by its adjacency matrix $A \equiv \{a_{ij}\}$, whose element a_{ij} is equal to 1 if link (i, j) is present and is 0 otherwise. Each node i is equipped with an UMT that describes the discovery process of the agent i [40]. We indicate the urn i at time t as $\mathcal{U}_i(t)$, while $\mathcal{S}_i(t)$ denotes the sequence of balls generated up to time t . Notice that $\mathcal{U}_i(t)$ is an unordered multiset of size $U_i(t) = |\mathcal{U}_i(t)|$, while $\mathcal{S}_i(t)$ is an ordered multiset of size $|\mathcal{S}_i(t)| = t$. Each urn i is characterized by two parameters, ρ_i and ν_i . As in the original UMT, the *reinforcement* parameter ρ_i accounts for the number of balls of the same color that are added to the urn i whenever a ball of a given color is extracted at time t . Furthermore, the *triggering* parameter ν_i controls the size of the *adjacent possible in the space of concepts*, as $(\nu_i + 1)$ balls of new colors are added to the urn of node i whenever at time t a color is extracted for the first time [40]. In this abstract representation, the space of concepts—made by all the colors—expands in time together with each discovery process, without relying on a predefined structure [41]. Discovery processes of different individuals are then coupled through the links of the network, representing social interactions. Namely, at each time t , the individual i draws a ball from an enriched urn, the so-called *social urn* of node i , $\tilde{\mathcal{U}}_i(t)$, composed by its own urn plus the additional balls present at time t in the urns of its neighbors, without their reinforcement. The latter represents the AP in the social space. Figure 1 illustrates the case of two nodes with a directed link. We thus have

$$\tilde{\mathcal{U}}_i(t) = \mathcal{U}_i(t) + \bigcup_{j \in \mathcal{N}} a_{ij} \mathcal{U}_j(t), \quad (1)$$

where $\mathcal{U}'_j(t) = \mathcal{U}_j^{[m=1]}(t) \subseteq \mathcal{U}_j(t)$ is the underlying set of the multiset $\mathcal{U}_j(t)$ (with multiplicity $m = 1$), i.e., the set of

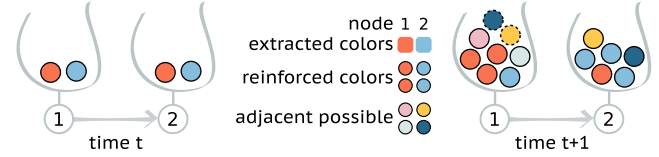


FIG. 1. Illustration of the model in the case of a network with two nodes. Each node is equipped with an urn obeying the UMT with the same parameters $\rho = 1$, $\nu = 1$, and $M_0 = \nu + 1$. At the time t , the urns start with two balls, one red (R) and the other blue (B). Then, each node extracts a ball (1, R ; 2, B), and, therefore, ρ additional balls of the same colors are added to the respective urns (reinforcement). Also, since in both cases, the extracted balls represent a novelty for the respective nodes, $\nu + 1$ balls of new colors are also added (adjacent possible). At $t + 1$, node 1 has access to all its balls plus two extra ones coming from the adjacent possible in the social space, i.e., the set of balls available through its neighbor (dashed borders).

size $U'_j(t) = |\mathcal{U}'_j(t)|$ formed by its unique elements. Duplicates in the urn associated to node j at time t are indeed not considered. Thus, the “memory” of node j due to the reinforcement does not influence node i . Similarly, let us denote with $\mathcal{S}'_i(t)$ the underlying set of the sequence $\mathcal{S}_i(t)$, i.e., the sequence of all the unique elements of $\mathcal{S}_i(t)$. We consider synchronous updates for all the urns.

Pace of discovery.—As previous works have shown [40], the dynamics of novelties and innovations share a number of commonalities and can, thus, be thought as two sides of the same process; a novelty refers to the discovery of something by an individual (already known to others), while innovations are novelties that are new to everybody. Here, we are interested in studying the asymptotic growth of the number of novelties—of each sequence—as a function of time (sequence length), representing the pace of discovery. We know, from standard results on the UMT [40], that an isolated urn i follows a Heaps’ law, i.e., a power law behavior $D_i(t) \sim t^{\beta_i}$ [44], $D_i(t) = |\mathcal{S}'_i(t)|$ being the number of different elements contained in the sequence $\mathcal{S}_i(t)$ up to time t . Thus, the Heaps’ exponent β_i quantifies the speed at which the urn discovers new elements (by definition bounded by $\beta_i \leq 1$). Let us consider now a node i that interacts through the network. In general, since $D_i(t)$ increases by one every time a ball is extracted for the first time, we can write $D_i(t + 1) = D_i(t) + P_i^{\text{new}}(t)$, where $P_i^{\text{new}}(t) \in [0, 1]$ is the probability that the ball extracted at node i at time t never appeared in $\mathcal{S}_i(t)$ before. In other words, $P_i^{\text{new}}(t) = \text{Prob}[D_i(t + 1) = D_i(t) + 1 | D_i(t)]$, and we can express it as the fraction of discoverable balls over the total number of balls available to node i at time t . This leads to an equation for the asymptotic Heaps’ dynamics that in the continuous time limit reads

$$\frac{dD_i(t)}{dt} = P_i^{\text{new}}(t) = \frac{|\tilde{\mathcal{U}}_i(t) \ominus \mathcal{S}'_i(t)|}{\tilde{U}_i(t)}, \quad (2)$$

where $\mathcal{A} \ominus \mathcal{B}$ denotes the multiset obtained by removing all the elements in set \mathcal{B} from the multiset \mathcal{A} (all duplicates are removed). Notice that if a node i has an out degree $\sum_j a_{ij} = 0$, its associated Eq. (2) reduces to the one of an isolated urn, for which $\tilde{U}_i(t) = U_i(t)$. Thus, its Heaps' dynamics for $\rho > \nu$ follows $D_i(t) \sim t^{\nu/\rho}$ for $t \rightarrow \infty$ [40,51] (see Supplemental Material [67]). In the most general case, where each node i is equipped with an UMT(ρ_i, ν_i), the equation for the Heaps' laws of each node $i \in \mathcal{N}$ can be written as in Eq. (2), by accounting for all the neighbors that are part of the social urn of node i . This can be done by using the nonzero elements of \mathbf{A} , so that the number of balls $\tilde{U}_i(t)$ in the social urn of node i at time t reads

$$\tilde{U}_i(t) = \rho_i t + \sum_{j \in \mathcal{N}} [a_{ij} + \delta_{ij}] [M_0 + (\nu_j + 1) D_j(t)], \quad (3)$$

where M_0 is the initial number of balls in each urn and δ_{ij} stands for the Kronecker delta. Finally, the large time behavior of the number of different elements $D_i(t)$ for each node i can be written as

$$\frac{dD_i(t)}{dt} = \frac{M_0 \sum_j (a_{ij} + \delta_{ij}) + \sum_j [\delta_{ij} \nu_j + a_{ij} (\nu_j + 1)] D_j(t)}{\rho_i t + \sum_j (a_{ij} + \delta_{ij}) [M_0 + (\nu_j + 1) D_j(t)]}. \quad (4)$$

Equation (4) forms a system of N coupled nonlinear ordinary differential equations, with initial conditions $D_i(0) = 0 \forall i \in \mathcal{N}$, that can be numerically integrated for any network topology $\{a_{ij}\}$.

Numerical results.—We start exploring the behavior of our model on the famous Zachary karate club network (ZKC) [74], where each node is equipped with an UMT($\rho = 6, \nu = 3$) with same parameters and initial conditions. We run different simulations and observe, for each node i , the average growth of the number of distinct elements $D_i(t)$ as a function of time. We then extract the values of the Heaps' exponents of each node as $\beta_i = \beta_i(T)$, where $\beta_i(t) = \ln D_i(t) / \ln t$ and $T = 10^4$. Figure 2 shows the nodes of the networks colored accordingly. Notice the higher pace of discovery displayed by the notoriously

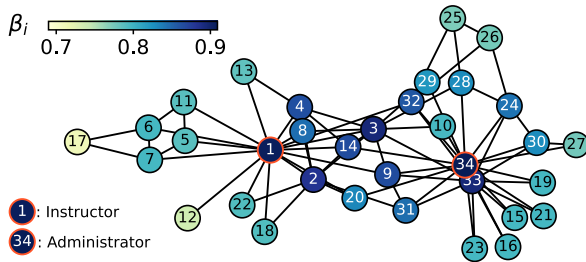


FIG. 2. Dynamics of the interacting urns on the Zachary karate club network, whose nodes are colored according to the resulting Heaps' exponent.

central nodes corresponding to the instructor (node 1) and the administrator of the club (node 34). This proves that nodes with identical UMTs can have completely different dynamics, suggesting that a strategic location on the social network correlates with the discovery potential of an individual. To further investigate this relation, we study the dynamics on five small directed networks. Figures 3(a)–3(e) show the temporal evolution of $D_i(t)$ for each node i of the networks displayed on the left. We report the simulated Heaps' laws (colored points), whose extracted exponents β_i are shown in the legend. In addition, to assess the validity of Eq. (4), we also plot the curves (continuous black lines) obtained using the appropriate $\{a_{ij}\}$. It can be seen that the analytical formalism introduced perfectly captures the Heaps' laws, since lines are almost indistinguishable from (simulated) points. In particular, in Fig. 3(a), we observe

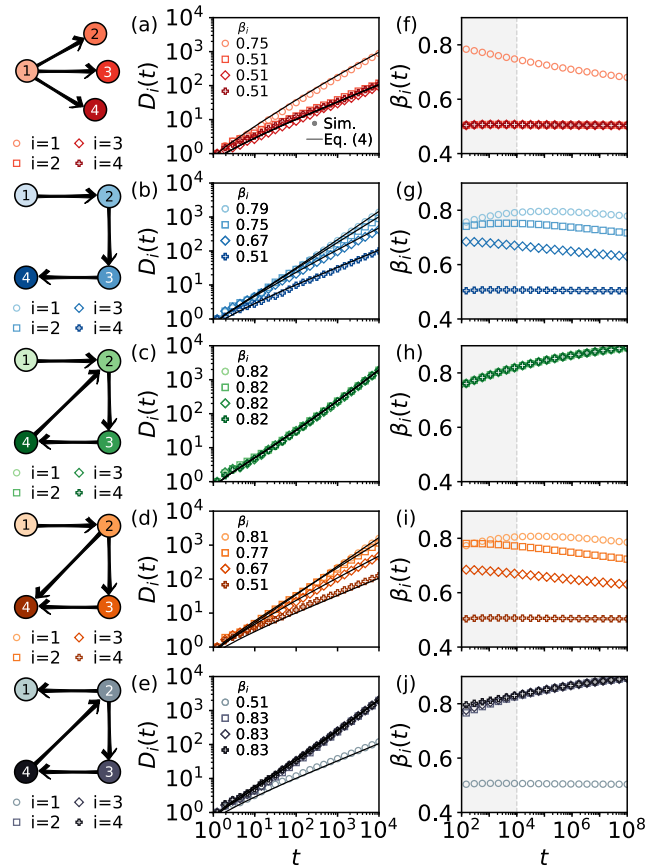


FIG. 3. Heaps' dynamics of the interacting urns on five directed toy graphs (different symbols correspond to different nodes). Each node is equipped with an UMT with the same parameters $\rho = 6$ and $\nu = 3$. (a)–(e) Temporal evolution of the number of discoveries $D_i(t)$ for each node i (associated Heaps' exponents β_i in the legend). The solutions of Eq. (4), shown as continuous black lines, are in perfect agreement with simulations. (f)–(j) Temporal behavior of the associated Heaps' exponents extracted at different times. The gray area up to $T = 10^4$ corresponds to the values of (a)–(e).

the highest pace of discovery in the node with more outgoing links. However, the nontrivial behaviors observed in Figs. 3(b)–3(e) for chains and graphs containing cycles indicate that the exponent of a node does not depend solely on local node properties. For instance, in Fig. 3(d), node 2 has two outgoing links, while the others have one link only. In contrast with what is observed in Fig. 3(a), here the highest pace of discovery is the one of node 1, whose social urn gets the benefits of the urn of node 2. Moreover, in Figs. 3(c) and 3(d), a simple change of direction of link $4 \rightarrow 2$ translates into completely different dynamics. We also notice that in both Figs. 3(c) and 3(e) the presence of a cycle enhances the pace of discovery in a process of mutual exchange. However, while in Fig. 3(d) node 1 is linked to the cycle and captures the same behavior of those in the cycle, in Fig. 3(e) node 1 behaves as an individual urn. We have further investigated whether the extracted β_i may depend on the maximum time T at which we have stopped the simulations. The curves reported in Figs. 3(f)–3(j) as a function of time for time up to 10^8 clearly indicate that the systems, even for the small graphs considered, have not yet reached a stationary state. Thermalization times, that are typical of empirical trajectories of diffusion process [75], here are strongly influenced by the topology of the network. This can be seen by comparing the two $\beta_1(t)$ of Figs. 3(f) and 3(g), both approaching—as we will see later—the asymptotic value $\nu/\rho = 0.5$ but at very different timescales. Nevertheless, the ranking induced by the pace of discovery persists at all finite times. In the next section, we will further investigate this characteristic behavior, ultimately proving its universality for all networks (see Supplemental Material [67]).

Analytical results.—In order to extract the asymptotic values of the Heaps' exponents and their dependence on the network topology, we derive an analytical solution of Eq. (4) for $t \rightarrow \infty$. Let us suppose $\rho_i = \rho$ and $\nu_i = \nu \forall i \in \mathcal{N}$. For sufficiently high values of ρ , we have $\lim_{t \rightarrow \infty} D_i(t)/t = 0 \forall i$, so that the denominator of the rhs of Eq. (4) can be approximated by ρt , leading to

$$\frac{d\vec{D}(t)}{dt} \approx \frac{1}{t} \left(\frac{\nu}{\rho} \mathbf{I} + \frac{\nu+1}{\rho} \mathbf{A} \right) \vec{D}(t) = \frac{1}{t} \mathbf{M} \vec{D}(t), \quad (5)$$

where $\vec{D}(t) \equiv \{D_i(t)\}_{i=1, \dots, N}$, \mathbf{I} denotes the $N \times N$ identity matrix, and we have introduced the constant matrix $\mathbf{M} = f(\mathbf{A}) = (\nu/\rho)\mathbf{I} + (\nu+1/\rho)\mathbf{A}$. By operating the change of variable $t = e^z$, Eq. (5) can be rewritten as $d_z \vec{D}(z) \approx \mathbf{M} \vec{D}(z)$, a standard first-order differential system, which leads to the solution

$$\vec{D}(t) \approx \sum_{\ell=1}^r \sum_{p=0}^{m_\ell-1} \vec{c}_p \ln^p(t) t^{\lambda_\ell}, \quad (6)$$

where $\{\lambda_\ell\}_{\ell=1, \dots, r}$ and $\{m_\ell\}_{\ell=1, \dots, r}$ are the eigenvalues of \mathbf{M} with their respective multiplicities and \vec{c}_p are vectors

defined by the initial conditions. The asymptotic behavior of $D_i(t)$ is then governed by the leading term in Eq. (6), so that

$$D_i(t) \underset{t \rightarrow \infty}{\approx} u_i \ln^{\hat{p}(i)}(t) t^{\hat{\lambda}(i)}, \quad (7)$$

where $\hat{\lambda}(i)$ is the eigenvalue of \mathbf{M} with the biggest real part such that the i th entry of at least one of its eigenvectors \vec{c}_p is different from zero. Similarly, $\hat{p}(i)$ is the maximum value of p among these eigenvectors and, in general, can be less than the multiplicity of the eigenvalue $\hat{\lambda}(i)$ minus one. For example, in the case of a chain as in Fig. 3(b), the asymptotic solution is $D_i(t) \sim u_i \ln^{N-i}(t) t^{\nu/\rho}$. In this example, all the exponents tend to ν/ρ at large times, while at finite times nodes with higher powers in the logarithm show higher paces of discovery, thus explaining the behavior seen in Fig. 3(g) (see Supplemental Material [67]).

In the case of strongly connected graphs, Eq. (7) simplifies: the logarithmic correction disappears, and all the asymptotic exponents are equal to the maximum eigenvalue $\hat{\lambda} = f(\hat{\mu})$ of \mathbf{M} . In fact, for the Perron-Frobenius theorem [76,77], \mathbf{A} has a simple and positive maximum eigenvalue $\hat{\mu}$ corresponding to an eigenvector \vec{u} with all positive entries. Thus, the approximated solution becomes

$$D_i(t) \underset{t \rightarrow \infty}{\approx} u_i t^{\hat{\lambda}}, \quad (8)$$

where u_i is proportional to the Bonacich eigenvector centrality [78] of node i , a global indicator of centrality that recursively quantifies the importance of a node from that of its neighbors and not just from the number of neighbors. As a consequence of Eq. (8), for strongly connected graphs, every node has approximately the same behavior $t^{\hat{\lambda}}$. What makes a node different from another is precisely the multiplicative factor u_i . In cycles and cliques, nodes are all structurally equivalent ($u_i = u \forall i$), meaning that they all have the same $D_i(t)$. On the contrary, in graphs such as the ZKC (see Fig. 2), the different values of u_i play a very important role. Most central nodes, as the instructor and the chief administrator, are the fastest explorers (highest β_i), even having the same asymptotic Heaps' exponent $\hat{\lambda}$.

In the more general case in which a graph is not strongly connected, Eq. (7) still holds, and the same argument can be applied to each of the strongly connected components to recursively find the values of u_i , $\hat{p}(i)$, and $\hat{\lambda}(i)$ (see Supplemental Material [67]). In such cases, the eigenvector centrality needs to be replaced by its natural extension to non-strongly-connected graphs, i.e., the α centrality [79]. We have investigated the correlation between the α centrality and the pace of discovery in real-world networks. Figure 4 shows the scatter plot of the number of discovered colors $D_i(T)$ and the normalized α centrality $c_i^{[\alpha]}/c_{\max}^{[\alpha]}$ in four empirical social networks: (a) the ZKC [74], (b) a

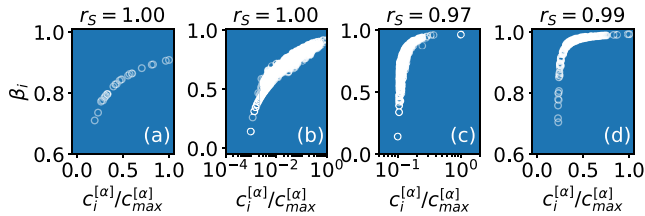


FIG. 4. Scatter plot and Spearman's rank correlation coefficients r_s between fitted Heaps' exponents β_i and normalized α centrality $c_i^{[\alpha]}/c_{\max}^{[\alpha]}$ associated to the $i = 1, \dots, N$ nodes of four empirical networks.

Twitter network of followers [80], (c) a coauthorship network in network science [81], and (d) a collaboration network between jazz musicians [82] (see Supplemental Material [67]). The high values of the Spearman's rank correlations ($r_s \geq 0.97$ in all cases) found in both undirected [Figs. 4(a), 4(c), and 4(d)] and directed networks [Fig. 4(b)] is in agreement with our predictions. This confirms that, together with the AP in the space of concepts, it is crucial to take into account of an AP in the social space.

In conclusion, we have presented a first example in which stochastic (and not deterministic) processes are coupled over the nodes of a complex network, and analytical insights on the relations between structure and dynamics are possible. The results highlight that the structural—not just local—properties of the nodes can strongly affect their ability to discover novelties. Our networked model of social urns is not just a simple extension of UMTs. What makes it novel and different is the very same idea of coupling together many urns over a complex social network and the concept of “social urn” we have introduced. It is such a network coupling that spontaneously produces novel behaviors, such as different exponents of the Heaps' law in a single system, and has the potential to open new areas of research and applications. This work represents only a first step toward the inclusion of structured interactions in discovery processes. Urns can, in fact, result in oversimplified models for the dynamics of individual explorers. Future works could consider non-identical urns or even explore the effects of having individuals with a finite storage capacity or where the adoption of the new might trigger the abandoning of the old, as for substitutive systems [83]. Another natural extension would be considering discoveries and social relationships unfolding across different network layers [84] or higher-order structures [85,86]. In addition, it would be interesting to study the relationship with existing models of social spreading and meme popularity [87–89]. Finally, our results could be directly applied in studies on efficient team structures in cooperative creative tasks [90–94].

I. I. and V. Latora acknowledge support from Engineering and Physical Sciences Research Council

(EPSRC) Grant No. EP/N013492/1. I. I. acknowledges support from the United Kingdom Regions Digital Research Facility (RDRF)-Urban Dynamics Lab under EPSRC Grant No. EP/M023583/1 and from The Alan Turing Institute under EPSRC Grant No. EP/N510129/1. V. L. acknowledges support from the Leverhulme Trust Research Fellowship 278 “CREATE: The network components of creativity and success.” G. D. B. thanks Sony CSL, where part of this work was completed, for the kind hospitality. We thank U. Alvarez-Rodriguez for the helpful comments and suggestions and V. D. P. Servedio and B. Monechi for interesting conversations about the early steps of this work.

I. I. and G. D. B. contributed equally to this work.

*i.iacopini@qmul.ac.uk

†v.latora@qmul.ac.uk

- [1] J. Drets, *Science* **287**, 1960 (2000).
- [2] D. H. Erwin and D. C. Krakauer, *Science* **304**, 1117 (2004).
- [3] F. Wu and B. A. Huberman, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 17599 (2007).
- [4] V. Sood, M. Mathieu, A. Shreim, P. Grassberger, and M. Paczuski, *Phys. Rev. Lett.* **105**, 178701 (2010).
- [5] M. Perc, *Sci. Rep.* **3**, 1720 (2013).
- [6] A. Rzhetsky, J. G. Foster, I. T. Foster, and J. A. Evans, *Proc. Natl. Acad. Sci. U.S.A.* **112**, 14569 (2015).
- [7] T. Fink, M. Reeves, R. Palma, and R. Farr, *Nat. Commun.* **8**, 2002 (2017).
- [8] A. T. Barron, J. Huang, R. L. Spang, and S. DeDeo, *Proc. Natl. Acad. Sci. U.S.A.* **115**, 4607 (2018).
- [9] M. Coccia, *Technol. Soc.* **59**, 1 (2019).
- [10] D. Park, J. Nam, and J. Park, *EPJ Data Sci.* **9**, 2 (2020).
- [11] B. Hofstra, V. V. Kulkarni, S. M.-N. Galvez, B. He, D. Jurafsky, and D. A. McFarland, *Proc. Natl. Acad. Sci. U.S.A.* **117**, 9284 (2020).
- [12] C. Cattuto, A. Barrat, A. Baldassarri, G. Schehr, and V. Loreto, *Proc. Natl. Acad. Sci. U.S.A.* **106**, 10511 (2009).
- [13] S. Thurner, P. Klimek, and R. Hanel, *New J. Phys.* **12**, 075029 (2010).
- [14] J. McNeerney, J. D. Farmer, S. Redner, and J. E. Trancik, *Proc. Natl. Acad. Sci. U.S.A.* **108**, 9008 (2011).
- [15] M. M. Dankulov, R. Melnik, and B. Tadić, *Sci. Rep.* **5**, 12197 (2015).
- [16] F. Saracco, R. Di Clemente, A. Gabrielli, and L. Pietronero, *PLoS One* **10**, e0140420 (2015).
- [17] M. Andjelković, B. Tadić, M. M. Dankulov, M. Rajković, and R. Melnik, *PLoS One* **11**, e0154655 (2016).
- [18] B. Tadić, M. M. Dankulov, and R. Melnik, *Phys. Rev. E* **96**, 032307 (2017).
- [19] T. M. A. Fink and M. Reeves, *Sci. Adv.* **5**, eaat6107 (2019).
- [20] T. M. A. Fink and A. Teimouri, *arXiv:1912.03281*.
- [21] M. Coccia, *Technological Forecasting and Social Change* **141**, 289 (2019).
- [22] A. Pichler, F. Lafond, and J. D. Farmer, *arXiv:2003.00580*.
- [23] E. Ubaldi, R. Burioni, V. Loreto, and F. Tria, *arXiv:2003.00989*.

- [24] R. Pemantle *et al.*, *Probab. Surv.* **4**, 1 (2007).
- [25] M. Launay and V. Limic, [arXiv:1207.5635](https://arxiv.org/abs/1207.5635).
- [26] G. Aletti, I. Crimaldi, A. Ghiglietti, *Bernoulli* **26**, 1098 (2020).
- [27] F. M. Hoppe, *J. Math. Biol.* **20**, 91 (1984).
- [28] G. Pólya, *Ann. Inst. Henri Poincaré* **1**, 117 (1930).
- [29] M. V. Simkin and V. P. Roychowdhury, *Phys. Rep.* **502**, 1 (2011).
- [30] M. Hayhoe, F. Alajaji, and B. Gharesifard, in *Proceedings of the American Control Conference (ACC), 2017* (IEEE, New York, 2017), pp. 358–363.
- [31] M. Hayhoe, F. Alajaji, and B. Gharesifard, in *Proceedings of the 2018 Annual American Control Conference (ACC)* (IEEE, New York, 2018), pp. 2644–2650.
- [32] S. Berg, *Public Choice* **47**, 377 (1985).
- [33] T. Gong, L. Shuai, M. Tamariz, and G. Jäger, *PLoS One* **7**, e33171 (2012).
- [34] C. Cattuto, V. Loreto, and L. Pietronero, *Proc. Natl. Acad. Sci. U.S.A.* **104**, 1461 (2007).
- [35] R. Marcaccioli and G. Livan, *Nat. Commun.* **10**, 745 (2019).
- [36] S. A. Kauffman, in *SFI Working Papers* (Santa Fe Institute, Santa Fe, NM, 1996).
- [37] N. H. Packard, *Dyn. Patterns Complex Syst.* **212**, 293 (1988).
- [38] C. Langton, Computation at the edge of chaos: Phase transition and emergent computation, Technical Report No. 1–3, 1990.
- [39] C. Langton, C. Taylor, J. Farmer, and S. Rasmussen, *Artificial Life II* (Avalon, New York, 2003).
- [40] F. Tria, V. Loreto, V. D. P. Servedio, and S. H. Strogatz, *Sci. Rep.* **4**, 5890 (2014).
- [41] V. Loreto, V. D. Servedio, S. H. Strogatz, and F. Tria, in *Creativity and Universality in Language* (Springer, New York, 2016), pp. 59–83.
- [42] P. Gravino, B. Monechi, V. Servedio, F. Tria, and V. Loreto, in *Proceedings of the Seventh International Conference on Computational Creativity, Paris, France* (2016), <http://www.computationalcreativity.net/iccc2016/wp-content/uploads/2016/01/Crossing-the-horizon.pdf>.
- [43] B. Monechi, Á. Ruiz-Serrano, F. Tria, and V. Loreto, *PLoS One* **12**, e0179303 (2017).
- [44] H. S. Heaps, *Information Retrieval: Computational and Theoretical Aspects* (Academic, New York, 1978).
- [45] G. K. Zipf, *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology* (Ravenio, Cambridge, England, 2016).
- [46] F. Font-Clos, G. Boleda, and A. Corral, *New J. Phys.* **15**, 093033 (2013).
- [47] M. Perc, *J. R. Soc. Interface* **11**, 20140378 (2014).
- [48] A. Mazzolini, A. Colliva, M. Caselle, and M. Osella, *Phys. Rev. E* **98**, 052139 (2018).
- [49] A. Mazzolini, M. Gherardi, M. Caselle, M. Cosentino Lagomarsino, and M. Osella, *Phys. Rev. X* **8**, 021023 (2018).
- [50] A. Mazzolini, J. Grilli, E. De Lazzari, M. Osella, M. C. Lagomarsino, and M. Gherardi, *Phys. Rev. E* **98**, 012315 (2018).
- [51] F. Tria, V. Loreto, and V. Servedio, *Entropy* **20**, 752 (2018).
- [52] I. Iacopini, S. Milojević, and V. Latora, *Phys. Rev. Lett.* **120**, 048301 (2018).
- [53] A. Mastrototaro, A mathematical model for the emergence of innovations, Ph.D. Thesis, Politecnico di Torino, 2018.
- [54] M. J. Salganik, P. S. Dodds, and D. J. Watts, *Science* **311**, 854 (2006).
- [55] R. Pálovics and A. A. Benczúr, *Soc. Network Anal. Mining* **5**, 4 (2015).
- [56] J. Ternovski and T. Yasseri, *Soc. Networks* **61**, 144 (2020).
- [57] P. F. Lazarsfeld, B. Berelson, and H. Gaudet, *The People's Choice* (Duell, Sloan & Pearce, New York, 1944).
- [58] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, *Nature (London)* **489**, 295 (2012).
- [59] J. Bryden, S. P. Wright, and V. A. Jansen, *J. R. Soc. Interface* **15**, 20170738 (2018).
- [60] R. Albert and A.-L. Barabási, *Rev. Mod. Phys.* **74**, 47 (2002).
- [61] M. E. Newman, *SIAM Rev.* **45**, 167 (2003).
- [62] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, Cambridge, England, 2017).
- [63] M. A. Porter and J. P. Gleeson, *Dynamical Systems on Networks: A Tutorial* (Springer, New York, 2005).
- [64] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical Processes on Complex Networks* (Cambridge University Press, Cambridge, England, 2008).
- [65] E. Rogers, *Diffusion of Innovations*, 4th ed. (Free Press, New York, 2010).
- [66] D. Centola, *How Behavior Spreads: The Science of Complex Contagions*, Princeton Analytical Sociology Series (Princeton University Press, Princeton, NJ, 2018).
- [67] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.125.248301> for the full derivation of the main equations and their analytical solution in the case of a single urn, two coupled urns, chains, cycles and cliques of N urns, and the general case of strongly and non-strongly-connected networks, for a description of the datasets, for a comprehensive investigation of the relationship between the ranking of the nodes induced by their Heaps' exponent and the eigenvector and α centrality, which includes Refs. [68–73].
- [68] R. Tarjan, *SIAM J. Comput.* **1**, 146 (1972).
- [69] L. Lü, D. Chen, X.-L. Ren, Q.-M. Zhang, Y.-C. Zhang, and T. Zhou, *Phys. Rep.* **650**, 1 (2016).
- [70] K. Ide, A. Namatame, L. Ponnambalam, F. Xiuju, and R. S. M. Goh, *Adv. Comput. Sci.* **3**, 115 (2014).
- [71] R. Ghosh and K. Lerman, *Phys. Rev. E* **83**, 066118 (2011).
- [72] R. Ghosh and K. Lerman, *Discrete Continuous Dyn. Syst. Ser. B* **19**, 1531 (2012).
- [73] L. Katz, *Psychometrika* **18**, 39 (1953).
- [74] W. W. Zachary, *J. Anthropological Res.* **33**, 452 (1977), www.jstor.org/stable/3629752.
- [75] G. Dosi, A. Moneta, and E. Stepanova, *Ind. Innovation* **26**, 461 (2019).
- [76] O. Perron, *Math. Ann.* **64**, 248 (1907).
- [77] G. Frobenius and S.-B. Deutsch. *Akad. Wiss. Berlin. Math-Nat. Kl.* 456 (1912).
- [78] P. Bonacich, *J. Math. Sociol.* **2**, 113 (1972).
- [79] P. Bonacich and P. Lloyd, *Soc. Networks* **23**, 191 (2001).

- [80] M. De Choudhury, Y.-R. Lin, H. Sundaram, K. S. Candan, L. Xie, and A. Kelliher, in *Proceedings of the Fourth International AAAI Conference on Weblogs and Social Media* (AAAI Press, 2010).
- [81] M. E. J. Newman, *Phys. Rev. E* **74**, 036104 (2006).
- [82] P. M. Gleiser and L. Danon, *Adv. Complex Syst.* **06**, 565 (2003).
- [83] C. Jin, C. Song, J. Bjelland, G. Canright, and D. Wang, *Nat. Hum. Behav.* **3**, 837 (2019).
- [84] S. Boccaletti, G. Bianconi, R. Criado, C. I. Del Genio, J. Gómez-Gardenes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin, *Phys. Rep.* **544**, 1 (2014).
- [85] I. Iacopini, G. Petri, A. Barrat, and V. Latora, *Nat. Commun.* **10**, 2485 (2019).
- [86] F. Battiston, G. Cencetti, I. Iacopini, V. Latora, M. Lucas, A. Patania, J.-G. Young, and G. Petri, *Phys. Rep.* **874**, 1 (2020).
- [87] J. P. Gleeson, J. A. Ward, K. P. O'Sullivan, and W. T. Lee, *Phys. Rev. Lett.* **112**, 048701 (2014).
- [88] J. P. Gleeson, K. P. O'Sullivan, R. A. Baños, and Y. Moreno, *Phys. Rev. X* **6**, 021019 (2016).
- [89] J. D O'Brien, I. K. Dassios, and J. P. Gleeson, *New J. Phys.* **21**, 025001 (2019).
- [90] A. Schecter, A. Pilny, A. Leung, M. S. Poole, and N. Contractor, *J. Organ. Behav.* **39**, 1163 (2018).
- [91] V. S. Torrisi, S. Manfredi, I. Iacopini, V. Latora *et al.*, in *DS 95: Proceedings of the 21st International Conference on Engineering and Product Design Education (E&PDE 2019)*, University of Strathclyde, Glasgow, 2019 (2019).
- [92] B. Monechi, G. Pullano, and V. Loreto, *Proc. Natl. Acad. Sci. U.S.A.* **116**, 22088 (2019).
- [93] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, *Science* **354**, aaf5239 (2016).
- [94] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi *et al.*, *Science* **359**, eaa0185 (2018).