Interacting Discovery Processes on Complex Networks

Iacopo Iacopini,[1, 2, 3] Gabriele Di Bona,[1, 4] Enrico Ubaldi,[5] 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, P.le Aldo Moro 5, 00185 Rome, Italy*
[7]*Complexity Science Hub Vienna (CSHV), Vienna, Austria*
[8]*Dipartimento di Fisica ed Astronomia, Università di Catania and INFN, I-95123 Catania, Italy*

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.

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 Pòlya 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* (S. Kauffman [36]). This formulation of the AP, which dates back to concepts previously introduced by D. Farmer, C. 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 sub-linear 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 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 multi-agent 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 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/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 two-fold 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/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 $\boldsymbol{A} \equiv \{a_{ij}\}$, 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 $U_i(t) = |\mathcal{U}_i(t)|$,
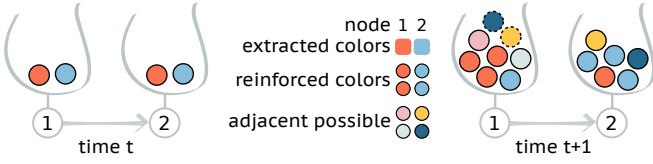
FIG. 1. Illustration of the model in the case of a network with two nodes. Each node is equipped with an urn obeying to the UMT with same parameters $\rho = 1$, $\nu = 1$, and $M_0 = \nu + 1$. At the time t, the urns start with two balls, one red (R) and the other blue (B). Then, each node extracts a 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).

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:

$$\tilde{\mathcal{U}}_i(t) = \mathcal{U}_i(t) + \bigcup_{j \in \mathcal{N}} a_{ij} \mathcal{U}'_j(t) \qquad (1)$$

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 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, $P_i^{\text{new}}(t) = \text{Prob}\left[D_i(t+1) = D_i(t) + 1 | D_i(t)\right]$ 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:

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

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 $\tilde{\mathcal{U}}_i(t) = \mathcal{U}_i(t)$. Thus, its Heaps dynamics for $\rho > \nu$ follows $D_i(t) \sim t^{\nu/\rho}$ for $t \to \infty$ [40, 51] (see Supplemental Material [67]). In the most general case, where each node i is equipped with a 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 non-zero elements of $\boldsymbol{A}$, so that the number of balls $\tilde{U}_i(t)$ in the social urn of node i at time t reads:

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

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

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

Eq.s (4) form a system of N coupled non-linear ODEs, 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 a 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
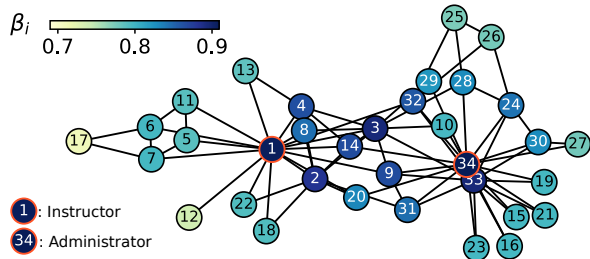
FIG. 2. Dynamics of the interacting urns on the Zachary Karate Club network, whose nodes are colored according to the resulting Heaps' exponent.

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 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. Figure 3(a-e) shows 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 the highest pace of discovery in the node with more outgoing links. However, the non-trivial behaviors observed in Fig. 3(b-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 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 Fig. 3(c) and (d) a simple change of direction of link $4 \rightarrow 2$ translates into completely different dynamics. We also notice that in both Fig. 3(c) and (e) the presence of a cycle enhances the pace of discovery in a process of mutual exchange. However, while in (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 Fig. 3(f-j) as 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 Fig. 3(f) and (g), both approaching—as we will see later—the asymptotic value $\nu/\rho = 0.5$ but at very different
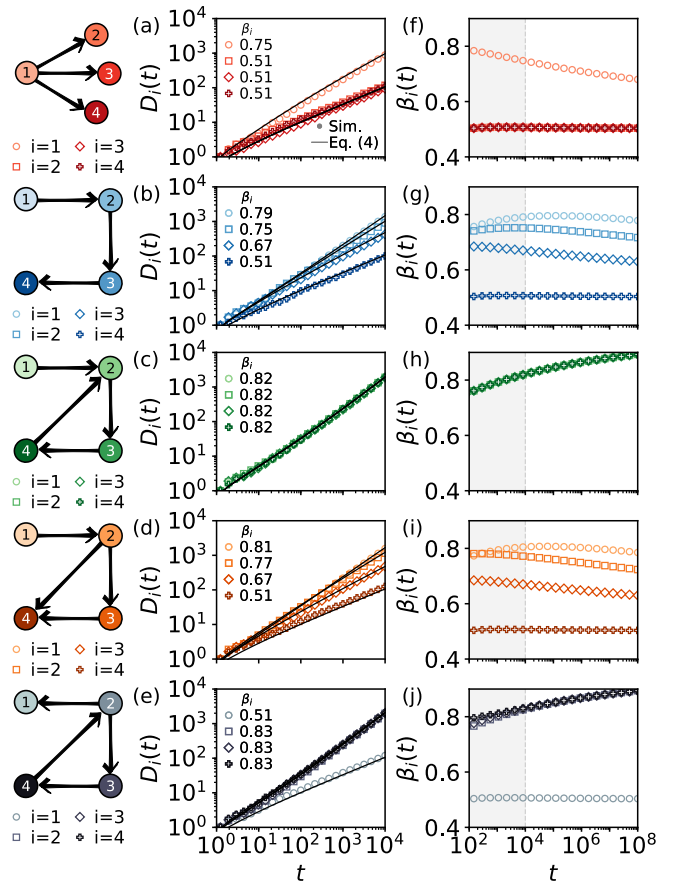

FIG. 3. Heaps' dynamics of the interacting urns on five directed toy graphs (different symbols correspond to different nodes). Each node is equipped with a UMT with same parameters $\rho = 6$, and $\nu = 3$. (a-e) Temporal evolution of the number of discoveries $D_i(t)$ for each node i (associated Heaps' 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 grey area up to $T = 10^4$ corresponds to the values of (a-e).

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 r.h.s. of Eq. (4) can be approximated by ρt, leading to:

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

where $\vec{D}(t) \equiv \{D_i(t)\}_{i=1,\ldots,N}$, $\boldsymbol{I}$ denotes the $N \times N$ identity matrix, and we have introduced the constant matrix $\boldsymbol{M} = f(\boldsymbol{A}) = \left(\frac{\nu}{\rho}\boldsymbol{I} + \frac{\nu+1}{\rho}\boldsymbol{A}\right)$. By operating the change of vari-
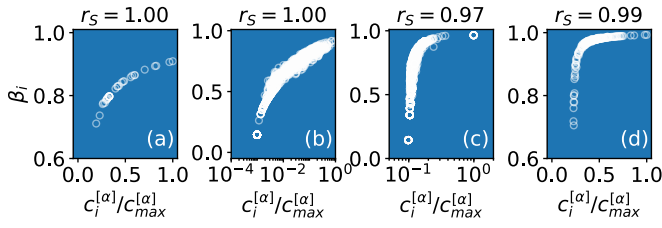
FIG. 4. Scatter plot and Spearman's rank correlation coefficients r_S between fitted Heaps' exponents β_i and normalized α-centrality $c_i^{[\alpha]}/c_{\max}^{[\alpha]}$ associated to the $i = 1, \ldots, N$ nodes of four empirical networks.

able $t = e^z$, Eq. (5) can be rewritten as $d_z \vec{D}(z) \approx \boldsymbol{M} \vec{D}(z)$, a standard first-order differential system, which leads to the solution

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

where $\{\lambda_\ell\}_{\ell=1,\ldots,r}$ and $\{m_\ell\}_{\ell=1,\ldots,r}$ are the eigenvalues of $\boldsymbol{M}$ 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:

$$D_i(t) \underset{t \to \infty}{\approx} u_i \ln^{\widehat{p}(i)}(t) \, t^{\widehat{\lambda}(i)}. \qquad (7)$$

where $\widehat{\lambda}(i)$ is the eigenvalue of $\boldsymbol{M}$ 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, $\widehat{p}(i)$ is the maximum value of p among these eigenvectors, and in general can be less than the multiplicity of the eigenvalue $\widehat{\lambda}(i)$ 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 $\widehat{\lambda} = f(\widehat{\mu})$ of $\boldsymbol{M}$. In fact, for the Perron-Frobenius theorem [76, 77], $\boldsymbol{A}$ has a simple and positive maximum eigenvalue $\widehat{\mu}$ corresponding to an eigenvector $\vec{u}$ with all positive entries. Thus, the approximated solution becomes:

$$D_i(t) \underset{t \to \infty}{\approx} u_i \, t^{\widehat{\lambda}}, \qquad (8)$$

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 $t^{\widehat{\lambda}}$. 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 $\widehat{\lambda}$.

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, $\widehat{p}(i)$, and $\widehat{\lambda}(i)$ (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 Twitter network of followers [80], (c) a co-authorship 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 (a,c,d) and directed networks (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 towards the inclusion of structured interactions in discovery processes. Urns can in fact result 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 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].

---

[1] J. Drews, 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**, 1 (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. in Soc. **59**, 101124 (2019).
[10] D. Park, J. Nam, and J. Park, EPJ Data Science **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. McNerney, 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** (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** (2019).
[20] T. M. A. Fink and A. Teimouri, arXiv:1912.03281 (2019).
[21] M. Coccia, Technol. Forecast. Soc. Change **141**, 289 (2019).
[22] A. Pichler, F. Lafond, and J. D. Farmer, arXiv:2003.00580 (2020).
[23] E. Ubaldi, R. Burioni, V. Loreto, and F. Tria, arXiv:2003.00989 (2020).
[24] R. Pemantle *et al.*, Probab. Surv **4**, 1 (2007).
[25] M. Launay and V. Limic, arXiv:1207.5635 (2012).
[26] G. Aletti, I. Crimaldi, A. Ghiglietti, *et al.*, Bernoulli **26**, 1098 (2020).
[27] F. M. Hoppe, J. Math. Biol. **20**, 91 (1984).
[28] G. Pólya, in *Annales de l'institut Henri Poincaré*, Vol. 1 (1930) pp. 117–161.
[29] M. V. Simkin and V. P. Roychowdhury, Phys. Rep. **502**, 1 (2011).
[30] M. Hayhoe, F. Alajaji, and B. Gharesifard, in *American Control Conference (ACC), 2017* (IEEE, 2017) pp. 358–363.
[31] M. Hayhoe, F. Alajaji, and B. Gharesifard, in *2018 Annual American Control Conference (ACC)* (IEEE, 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. Comm. **10**, 745 (2019).
[36] S. A. Kauffman, in *SFI working papers* (Santa Fe Institute, 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*, Tech. Rep. 1–3 (1990).
[39] C. Langton, C. Taylor, J. Farmer, and S. Rasmussen, *Artificial Life II* (Avalon Publishing, 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, 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* (2016).
[43] B. Monechi, Ã. Ruiz-Serrano, F. Tria, and V. Loreto, PLoS One **12**, e0179303 (2017).
[44] H. S. Heaps, *Information retrieval: Computational and theoretical aspects* (Academic Press, Inc., 1978).
[45] G. K. Zipf, *Human behavior and the principle of least effort: An introduction to human ecology* (Ravenio Books, 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, Social Network Analysis and Mining **5**, 4 (2015).
[56] J. Ternovski and T. Yasseri, Soc. Netw. **61**, 144 (2020).
[57] P. F. Lazarsfeld, B. Berelson, and H. Gaudet, *The people's choice.* (Duell, Sloan & Pearce, 1944).
[58] R. M. Bond, C. J. Fariss, J. J. Jones, A. D. Kramer, C. Marlow, J. E. Settle, and J. H. Fowler, Nature **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 review **45**, 167 (2003).
[62] V. Latora, V. Nicosia, and G. Russo, *Complex Networks: Principles, Methods and Applications* (Cambridge University Press, 2017).
[63] M. A. Porter and J. P. Gleeson, *Dynamical systems on networks: A tutorial* (Springer, 2005).
[64] A. Barrat, M. Barthelemy, and A. Vespignani, *Dynamical pro-*

cesses on complex networks (Cambridge university press, 2008).

[65] E. Rogers, *Diffusion of Innovations, 4th Edition* (Free Press, 2010).

[66] D. Centola, *How Behavior Spreads: The Science of Complex Contagions*, Princeton Analytical Sociology Series (Princeton University Press, 2018).

[67] See Supplemental Material 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 adescription of the data sets, for a comprehensive investigation of the relationship between the ranking of the nodes induced by their Heap's exponent and the eigenvector and α-centrality, which includes Refs. [68-73],.

[68] R. Tarjan, SIAM journal on computing **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] K. Ghosh, Rumi; Lerman, Phys. Rev. E **83** (2011).

[72] R. Ghosh and K. Lerman, Discrete Continuous Dyn. Syst. Ser. B **19** (2012).

[73] L. Katz, Psychometrika **18**, 39 (1953).

[74] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[75] G. Dosi, A. Moneta, and E. Stepanova, Industry and Innovation **26**, 461 (2019).

[76] O. Perron, Math. Ann. **64**, 248 (1907).

[77] G. Frobenius, in *S.-B.Deutsch.Akad. Wiss. Berlin. Math-Nat. Kl.* (1912) pp. 456–477.

[78] P. Bonacich, J. Math. Sociol. **2**, 113 (1972).

[79] P. Bonacich and P. Lloyd, Soc. Netw **23**, 191 (2001).

[80] M. De Choudhury, Y.-R. Lin, H. Sundaram, K. S. Candan, L. Xie, and A. Kelliher, in *Fourth International AAAI Conference on Weblogs and Social Media* (2010).

[81] M. E. J. Newman, Phys. Rev. E **74**, 036104 (2006).

[82] P. M. Gleiser and L. Danon, Adv. Complex Syst. **6**, 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, Physics Reports **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. 12th-13th September 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**, eaao0185 (2018).

[95] A. Berman and R. J. Plemmons, "Nonnegative matrices," in *Nonnegative Matrices in the Mathematical Sciences* (Academic Press, 1979) Chap. 2, pp. 26–62.

Supplemental Material: Interacting Discovery Processes on Complex Networks

I. ANALYTICAL SOLUTIONS

In this section, we will study in more detail the analytical solutions we derived in the main text. We start by reviewing the case of an individual urn, which is equivalent to the urn model with triggering. We will then move on to more complicated cases, such as a pair of nodes, a chain, a cycle, a clique, ending with the formulation for the very general networks. Moreover, we will derive an algorithmic solution that allows deriving an analytical solution for each of the small networks studied in Fig. 3 of the main text. In every case, we will set the same parameters for each urn, so that $\rho_i = \rho$ (*reinforcement*) and $\nu_i = \nu$ (*triggering*) $\forall i = 1, \ldots, N$. Each urn will be initialized with M_0 balls of different colors. These and the other colors—added from an individual i when triggered by a discovery—will be taken from a single predefined set of discoverable balls of different colors. Notice that this set is shared by all the urns so that once a ball is drawn from an urn, it will not be available anymore to the others, except when enlarging the urn through the social adjacent possible (if they are connected).

A. The single urn model

Let us consider the simplest case of an isolated urn, or equivalently, an urn on a node i for which the out-degree $\sum_j a_{ij}$ is null. In this case, the dynamics will be the same of the Urn Model with Triggering (UMT) [40, 51], since the node does not have access to the balls of the neighbors, implying that its social urn will not be enriched ($\tilde{\mathcal{U}}_i(t) = \mathcal{U}_i(t)$). For such a node, the equation for the asymptotic Heaps' dynamics, Eq. (2) of the main text, reduces to:

$$\frac{dD(t)}{dt} = \frac{|\tilde{\mathcal{U}}(t) \ominus \mathcal{S}'(t)|}{\tilde{U}(t)} = \frac{U'(t) - D(t)}{U(t)}, \tag{S1}$$

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). Equation (S1) can now be written as a function of the parameters of the model. In particular, we can write the total number of balls in the urn up to time t, $U(t)$, as the initial number of balls M_0, plus the ρ balls added (t times) as reinforcement, plus the $(\nu + 1)$ balls added ($D(t)$ times, one for each novelty) due to the triggering mechanism:

$$U(t) = M_0 + \rho t + (\nu + 1)D(t). \tag{S2}$$

Similarly, the number of unique elements in the urn at time t, $U'(t)$, can be obtained by subtracting from $U(t)$ the ρt repeated balls coming from the reinforcement, that is:

$$U'(t) - D(t) = [U(t) - \rho t] - D(t) = M_0 + \nu D(t). \tag{S3}$$

Thus, using Eq. (S2) and Eq. (S3) in Eq. (S1) we obtain:

$$\frac{dD(t)}{dt} = \frac{M_0 + \nu D(t)}{M_0 + \rho t + (\nu + 1)D(t)}. \tag{S4}$$

From now onwards we suppose that $t \gg M_0$, so that we can disregard M_0 in Eq. (S4) and in the similar equations we will obtain in the following sections. Therefore, after the introduction of the auxiliary variable $z(t) = \frac{D(t)}{t}$, Eq. (S4) can be rewritten as:

$$\frac{dz(t)}{dt}t + z(t) = \frac{\nu z(t)t}{\rho t + (\nu + 1)z(t)t}, \tag{S5}$$

which can be integrated as:

$$\int_{z(t_0)}^{z(t)} \frac{\rho + (\nu + 1)z(t)}{z(t)[\nu - (\nu + 1)z(t) - \rho]}dz(t) = \int_{t_0}^{t} \frac{1}{t}dt. \tag{S6}$$

The asymptotic solution ($t \to \infty$) depends on the parameters ρ and ν. It can be shown, as in the Supplemental Material of Ref. [40, 51], that the asymptotic solution for $D(t)$ is

$$\begin{cases} \rho > \nu & D(t) \sim (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}} \\ \rho = \nu & D(t) \sim \frac{\nu}{\nu+1} \frac{t}{\ln t} \\ \rho < \nu & D(t) \sim \frac{\nu - \rho}{\nu + 1}t \end{cases} \tag{S7}$$

that is precisely the Heaps' law [44], with sublinear growth for $\rho > \nu$ and linear for the other cases. As empirical data has shown [40, 44], Heaps' laws usually have a sublinear behavior. For this reason, in this paper, we focus only on the case $\rho > \nu$.

B. Two coupled urns

Let us consider now the simplest case of two coupled urns, that is a network with only two nodes connected by a directed edge $(1 \to 2)$, as in Fig. 1 of the main text. This is equivalent to a directed chain of $N = 2$ nodes, that will be discussed in the next Section for a general number N of nodes. The associated equations to determine the asymptotic Heaps' laws can be written expressing the probabilities $P_i^{\text{new}}(t)$ to draw a new ball as the the fraction of discoverable balls over the total number of balls available to node i at time t:

$$
\begin{cases}
\dfrac{dD_1(t)}{dt} = \dfrac{|\tilde{\mathcal{U}}_1(t) \ominus \mathcal{S}_1'(t)|}{\tilde{U}_1(t)} & \text{(S8a)} \\[4mm]
\dfrac{dD_2(t)}{dt} = \dfrac{|\tilde{\mathcal{U}}_2(t) \ominus \mathcal{S}_2'(t)|}{\tilde{U}_2(t)} = \dfrac{U_2'(t) - D_2(t)}{U_2(t)}. & \text{(S8b)}
\end{cases}
$$

Notice that the right-hand side of Eq. (S8b) is simplified since node 2 does not have any outgoing link, and therefore its dynamics is the same of an isolated urn for which $\tilde{\mathcal{U}}_2(t) = \mathcal{U}_2(t)$. Thus, following the procedure discussed in the previous section, we have, for $\rho > \nu$:

$$
D_2(t) \sim (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}. \tag{S9}
$$

The denominator $\tilde{U}_1(t)$ of Eq. (S8a) can be expressed in terms of the two contributions coming from the two urns at time t, which reads:

$$
\tilde{U}_1(t) = \overbrace{M_0 + \rho t + (\nu + 1)D_1(t)}^{U_1(t)} + \overbrace{M_0 + (\nu + 1)D_2(t)}^{U_2'(t)} \tag{S10}
$$
$$
= 2M_0 + \rho t + (\nu + 1)\big[D_1(t) + D_2(t)\big].
$$

Similarly, the numerator of Eq. (S8a), consisting in the number of balls present in the social urn of node 1 at time t which did not appeared yet in $\mathcal{S}_1(t)$, can be written as the total number of balls in the social urn of 1 at time t, minus the number of duplicates, minus the number of balls that do not represent a novelty anymore with respect to the sequence $\mathcal{S}_1(t)$, i.e.:

$$
|\tilde{\mathcal{U}}_1(t) \ominus \mathcal{S}_1'(t)| = \tilde{U}_1(t) - \rho t - D_1(t). \tag{S11}
$$

Then, using Eq. (S10) and Eq. (S11), the final expression for Eq. (S8a) reads:

$$
\frac{dD_1(t)}{dt} = \frac{2M_0 + \nu D_1(t) + (\nu + 1)D_2(t)}{2M_0 + \rho t + (\nu + 1)\big[D_1(t) + D_2(t)\big]}. \tag{S12}
$$

For large times ($t \gg M_0$) we can approximate Eq. (S12) as

$$
\frac{dD_1(t)}{dt} \approx \frac{\nu D_1(t) + (\nu + 1)D_2(t)}{\rho t + (\nu + 1)\big[D_1(t) + D_2(t)\big]}. \tag{S13}
$$

Let us assume now that the dynamics of node 2 relaxes before the one of node 1, so that we can solve Eq. (S13) independently from Eq. (S9). In addition, if we suppose that $\lim_{t \to \infty} D(t)/t = 0$, Eq. (S13) can be approximated as:

$$
\frac{dD_1(t)}{dt} \approx \frac{\nu D_1(t)}{\rho t} + \frac{(\nu + 1)D_2(t)}{\rho t}. \tag{S14}
$$

The related homogeneous equation has a similar solution of Eq. (S9), i.e.:

$$
\frac{d\overline{D}_1(t)}{dt} \approx \frac{\nu \overline{D}_1(t)}{\rho t} \implies \overline{D}_1(t) \sim (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}. \tag{S15}
$$

We now look for a solution like $D_1(t) = \kappa(t)\overline{D}_1(t)$ that, plugged into Eq. (S14), leads to:

$$\frac{d\kappa(t)}{dt}\overline{D}_1(t) + \kappa(t)\frac{d\overline{D}_1(t)}{dt} \approx \kappa(t)\frac{d\overline{D}_1(t)}{dt} + \frac{(\nu+1)D_2(t)}{\rho t}. \tag{S16}$$

Thus, from Eq. (S9) and Eq. (S15) we get:

$$\frac{d\kappa(t)}{dt} = \frac{\nu+1}{\rho t}\frac{D_2(t)}{\overline{D}_1(t)} \approx \frac{\nu+1}{\rho t}, \tag{S17}$$

whose solution is

$$\kappa(t) \approx \frac{\nu+1}{\rho}\ln t. \tag{S18}$$

The asymptotic solution ($t \to \infty$) of $D_1(t)$ is then approximated by:

$$D_1(t) \sim \frac{\nu+1}{\rho}(\rho-\nu)^{\frac{\nu}{\rho}}\ln(t)\,t^{\frac{\nu}{\rho}}. \tag{S19}$$

In conclusion, comparing the solutions in Eq. (S9) and Eq. (S19) the presence of an outgoing link increases the number of novelties with respect to an isolated urn dynamics. However, as we have shown here, this increase is approximately only logarithmic, meaning that we can see a slight increase at finite times which practically disappears for larger times. Le us also notice that this applies to the directed case, while in the case of an undirected link we would get identical Heaps' laws for both nodes $i = 1, 2$, without logarithmic corrections, but with higher exponents. This particular case is a cycle of two nodes, and as we will see in a dedicated section, cycles have their own behavior.

C. Chain of N urns

Let us consider now a slightly more complicated case. Let us suppose that the network is composed by an open chain of N urns, where there are only directed links $(i \to i+1)$, with $i = 1, 2, \ldots, N-1$. This is the case considered in Fig. 3(b, g) of the main text, where in that case $N = 4$. Analogously to the previous case, the associated set of equations governing the growth of the number of novelties can be approximated to:

$$\begin{cases} \dfrac{dD_1(t)}{dt} \approx \dfrac{\nu D_1(t) + (\nu+1)D_2(t)}{\rho t + (\nu+1)\big[D_1(t) + D_2(t)\big]} & \text{(S20a)} \\[4mm] \qquad\vdots & \text{(S20b)} \\[2mm] \dfrac{dD_{N-1}(t)}{dt} \approx \dfrac{\nu D_{N-1}(t) + (\nu+1)D_N(t)}{\rho t + (\nu+1)\big[D_{N-1}(t) + D_N(t)\big]} & \text{(S20c)} \\[4mm] \dfrac{dD_N(t)}{dt} \approx \dfrac{\nu D_N(t)}{\rho t + (\nu+1)D_N(t)} & \text{(S20d)} \end{cases}$$

We can solve the system by solving each equation, starting from the last one and recursively substituting its solution into the equation above. Indeed, since node $i = N$ does not have any outgoing link its independent Eq. (S20d) can be immediately solved, resulting in the known asymptotic solution:

$$D_N(t) \sim (\rho-\nu)^{\frac{\nu}{\rho}}t^{\frac{\nu}{\rho}}. \tag{S21}$$

As in the previous case, in Eq. (S20c) we can consider $D_{N-1}(t)$ to be the only unknown variable. Then, following the same analytical steps presented in previous section leads to:

$$D_{N-1}(t) \approx \frac{\nu+1}{\rho}(\rho-\nu)^{\frac{\nu}{\rho}}\ln(t)t^{\nu/\rho}. \tag{S22}$$

The same reasoning can be iterated for each node i. Let us now prove by induction on i that the asymptotic solution is

$$D_i(t) = \frac{(\rho-\nu)^{\nu/\rho}}{(N-i)!}\left(\frac{\nu+1}{\rho}\ln(t)\right)^{N-i}t^{\nu/\rho}. \tag{S23}$$

We have already proved that this holds for $i = N$ and $i = N - 1$. Let us now suppose that it holds for i and let us prove it for $i - 1$, with $1 < i < N$. In the asymptotic limit, the equation for the growth of the number of novelties of node i reads

$$\frac{dD_{i-1}(t)}{dt} \approx \frac{\nu D_{i-1}(t) + (\nu + 1)D_i(t)}{\rho t + (\nu + 1)\big[D_{i-1}(t) + D_i(t)\big]}. \tag{S24}$$

For the induction hypothesis, in Eq. (S24) the only unknown variable is $D_i(t)$. Therefore, we can consider the homogeneous associated equation

$$\frac{d\overline{D}_{i-1}(t)}{dt} \approx \frac{\nu \overline{D}_{i-1}(t)}{\rho t}, \tag{S25}$$

which provides the approximated solution:

$$\overline{D}_{i-1}(t) \approx (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}. \tag{S26}$$

As for the case of two coupled urns, we now look for a solution like $D_{i-1}(t) = \kappa(t)\overline{D}_{i-1}(t)$, that, plugged into Eq. (S24), leads to:

$$\frac{d\kappa(t)}{dt}\overline{D}_{i-1}(t) + \kappa(t)\frac{d\overline{D}_{i-1}(t)}{dt} \approx \kappa(t)\frac{d\overline{D}_{i-1}(t)}{dt} + \frac{(\nu + 1)D_i(t)}{\rho t}. \tag{S27}$$

Thus, we get

$$\frac{d\kappa(t)}{dt} \approx \frac{\nu + 1}{\rho t}\frac{D_i(t)}{\overline{D}_{i-1}(t)} \approx \frac{1}{(N - i)!}\frac{\nu + 1}{\rho t}\left(\frac{\nu + 1}{\rho}\ln(t)\right)^{N - i}, \tag{S28}$$

whose solution is

$$\kappa(t) \approx \frac{1}{(N - (i - 1))!}\left(\frac{\nu + 1}{\rho}\ln(t)\right)^{N - (i - 1)}. \tag{S29}$$

Finally, after combining Eq. (S26) and Eq. (S29), we reach the solution for the dynamics of node $i - 1$, that reads:

$$D_{i-1}(t) \approx \frac{(\rho - \nu)^{\nu/\rho}}{(N - (i - 1))!}\left(\frac{\nu + 1}{\rho}\ln(t)\right)^{N - (i - 1)} t^{\nu/\rho}, \tag{S30}$$

which completes the proof by induction.

Finally, it is worth observing that the Heaps' laws would be very different if the links were undirected. This would indeed result, similarly to undirected cycles, in higher asymptotic Heaps' exponents.

D. Cycle of N urns

Directed cycle—Let us consider the case of directed cycles. As we will see, this is the simplest system leading to asymptotic Heaps' exponents that are higher than that of an individual urn. Let us hence suppose that every node i is connected just to the following one, node $i + 1$, with a directed link $(i \to i + 1)$, with $i = 1, 2, \ldots, N$, where we identify node $N + 1$ with node 1. For a generic node i, the asymptotic differential equation for the growth of the number of novelties reads:

$$\frac{dD_i(t)}{dt} \approx \frac{\nu D_i(t) + (\nu + 1)D_{i+1}(t)}{\rho t + (\nu + 1)\big[D_i(t) + D_{i+1}(t)\big]}. \tag{S31}$$

For symmetry reasons, the dynamics of each node is the same, implying that $D_1(t) \approx \cdots \approx D_i(t) \approx \cdots \approx D_N(t)$. Hence, Eq. (S31) becomes

$$\frac{dD_i(t)}{dt} \approx \frac{(2\nu + 1)D_i(t)}{\rho t + 2(\nu + 1)D_i(t)}, \tag{S32}$$

that is equal to the equation of an individual urn [see Eq. (S4)], with $\nu' = 2\nu + 1$. Therefore, if $\rho > \nu'$ we have the solution

$$D_i(t) \approx (\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{2\nu+1}{\rho}}. \tag{S33}$$

Undirected cycle—Let us now consider cycles composed by undirected links. Let us suppose that $N > 2$, considered that for $N = 1$ the network reduces to an individual urn, and for $N = 2$ it is equivalent to a directed cycle of 2 nodes. For $N > 2$, each node i is therefore connected to two different nodes $i - 1$ and $i + 1$, and the associated equations to be solved are:

$$\frac{dD_i(t)}{dt} \approx \frac{\nu D_i(t) + (\nu+1)D_{i-1}(t) + (\nu+1)D_{i+1}(t)}{\rho t + (\nu+1)\big[D_i(t) + (\nu+1)D_{i-1}(t) + D_{i+1}(t)\big]}. \tag{S34}$$

Again, for symmetry reasons, we can equivalently write Eq. (S34) as

$$\frac{dD_i(t)}{dt} \approx \frac{(3\nu+2)D_i(t)}{\rho t + 3(\nu+1)D_i(t)}, \tag{S35}$$

that is equal to the equation of an individual urn [see Eq. (S4)], with $\nu'' = 3\nu + 2$. Therefore, if $\rho > \nu''$ we have the solution

$$D_i(t) \approx (\rho - 3\nu - 2)^{\frac{3\nu+2}{\rho}} t^{\frac{3\nu+2}{\rho}}. \tag{S36}$$

Notice that for undirected cycles, since all connections are mutual, the resulting paces of discovery are higher than those in the directed case. However, in both cases, directed and undirected, the dynamics of each node does not depend on the length of the cycle.

E. Clique of N urns

Let us consider a N-clique, that is a fully connected network of N nodes, equivalently directed or undirected. Being every node i connected to all other nodes, all nodes are equivalent, and the general equation for the growth of the number of novelties of node i reads:

$$\frac{dD_i(t)}{dt} \approx \frac{\nu D_i(t) + (\nu+1)\sum_{j\neq i} D_j(t)}{\rho t + (\nu+1)\sum_{j=1}^{N} D_j(t)}. \tag{S37}$$

For symmetry reasons, each urn follows the same dynamics and we can equivalently write Eq. (S37) as

$$\frac{dD_i(t)}{dt} \approx \frac{[N(\nu+1)-1]D_i(t)}{\rho t + N(\nu+1)D_i(t)}, \tag{S38}$$

that is equal to the equation for an individual urn [see Eq. (S4)], with $\nu''' = N(\nu+1) - 1$. Therefore, if $\rho > \nu'''$ we have the solution

$$D_i(t) \approx (\rho - N(\nu+1) - 1)^{\frac{N(\nu+1)-1}{\rho}} t^{\frac{N(\nu+1)-1}{\rho}}. \tag{S39}$$

Let us observe that for any network with N nodes, the maximum allowed Heaps' exponent is hence $[N(\nu+1)-1]/\rho$, which occurs only in the case of a fully connected network.

F. The general case

Let us consider a general graph $G(\mathcal{N}, \mathcal{E})$, either directed or undirected. In order to write and solve the equations for the growth of the number of novelties, we first have to calculate the probability $P_i^{\text{new}}(t)$ of drawing a new ball from the urn of each node i. This can be done by considering the number of different colors present in the social urn $\tilde{\mathcal{U}}_i(t)$ of node i at time t that have not been discovered yet by i, divided by the total number of balls $\tilde{U}_i(t)$ present in its social urn at that time. The numerator can be expressed as $|\tilde{\mathcal{U}}_i(t) \ominus \mathcal{S}'_i(t)|$, which is the length of the multiset obtained by removing from the multiset $\tilde{\mathcal{U}}_i(t)$ all the elements appeared in the sequence (taking out all duplicates). In other words, it is the number of unique colors present in the urn of node i and in the one of its neighbors (without their multiplicity) minus the number of colors already drawn (unique elements in the

sequence of i). Considering that all (and only) the already discovered balls are those that have been reinforced and that the number of triggered colors added to the urn j is exactly $(\nu + 1)D_j(t)$, we can write:

$$\frac{dD_i(t)}{dt} = P_i^{\text{new}}(t) = \frac{|\tilde{\mathcal{U}}_i(t) \ominus \mathcal{S}'_i(t)|}{\tilde{U}_i(t)} = \frac{M_0 + \nu D_i(t) + \sum_{j \neq i} a_{ij}\big[M_0 + (\nu + 1)D_j(t)\big]}{\rho t + M_0 + (\nu + 1)D_i(t) + \sum_{j \neq i} a_{ij}\big[M_0 + (\nu + 1)D_j(t)\big]}, \tag{S40}$$

or, equivalently:

$$\frac{dD_i(t)}{dt} = \frac{M_0 \sum_j (a_{ij} + \delta_{ij}) + \sum_j \big[\delta_{ij}\nu + a_{ij}(\nu + 1)\big]D_j(t)}{\rho t + \sum_j (a_{ij} + \delta_{ij})\big[M_0 + (\nu + 1)D_j(t)\big]}. \tag{S41}$$

For $t \gg M_0$ we can disregard the presence of M_0 in Eq. (S41). As shown above for N-cliques, in the asymptotic limit $t \to \infty$ the growth of the number of novelties obeys an Heaps' law with maximum exponent $[N(\nu + 1) - 1]/\rho$. This means that if ρ is high enough, we can approximate the denominator on the r.h.s. of Eq. (S41) to ρt. After finding the approximated solution, we will estimate the set of parameters for which this approximation is valid for any topology. Therefore, in the asymptotic limit and with a proper choice of the parameters, Eq. (S41) can be rewritten as:

$$\frac{dD_i(t)}{dt} \approx \frac{\sum_j \big[\delta_{ij}\nu + a_{ij}(\nu + 1)\big]D_j(t)}{\rho t}, \tag{S42}$$

which can be expressed in a more compact way as:

$$\frac{d\vec{D}(t)}{dt} \approx \frac{1}{t}\left(\frac{\nu}{\rho}\boldsymbol{I} + \frac{\nu + 1}{\rho}\boldsymbol{A}\right)\vec{D}(t) = \frac{1}{t}\frac{f(\boldsymbol{A})\vec{D}(t)}{t} = \frac{1}{t}\boldsymbol{M}\vec{D}(t), \tag{S43}$$

where $\boldsymbol{I}$ is the $N \times N$ identity matrix and $\boldsymbol{M} = f(A)$, with $f(x) = \frac{\nu}{\rho} + \frac{\nu+1}{\rho}x$. By operating the change of variable $t = e^z$, Eq. (S43) can be rewritten as a standard first-order differential system, i.e. $d_z\vec{D}(z) \approx \boldsymbol{M}\vec{D}(z)$, which leads to the solution

$$\vec{D}(t) \approx \sum_{\ell=1}^{r} \sum_{p=0}^{m_\ell - 1} \vec{c}_p \ln^p(t)\, t^{\lambda_\ell}, \tag{S44}$$

where $\{\lambda_\ell\}_{\ell=1,\dots,r}$ and $\{m_\ell\}_{\ell=1,\dots,r}$ are the eigenvalues of $\boldsymbol{M}$ with their respective multiplicities, and $\vec{c}_p$ are vectors defined by the initial conditions. The asymptotic behavior of the number of novelties $D_i(t)$ discovered by node i at time t is then governed by the leading term in Eq. (S44), so that we can write:

$$D_i(t) \underset{t \to \infty}{\approx} u_i \ln^{\widehat{p}(i)}(t)\, t^{\widehat{\lambda}(i)}. \tag{S45}$$

where $\widehat{\lambda}(i)$ is the eigenvalue of $\boldsymbol{M}$ 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, $\widehat{p}(i)$ is the maximum value of p among these eigenvectors with non-zero i-th entries. In general, then, $\widehat{\lambda}(i)$ might not be the maximum eigenvalue of $\boldsymbol{M}$, like $\widehat{p}(i)$ might be less than the multiplicity of the eigenvalue $\widehat{\lambda}(i)$ minus one. Moreover, different nodes may have different values for these exponents. In particular, we have the same exponents for nodes in the same strongly connected components (SCCs), while they may vary from SCC to SCC. In the following paragraphs we will investigate this aspect.

Strongly connected network—Let us suppose that the graph $G(\mathcal{N}, \mathcal{E})$ is strongly connected. In this case the solution given by Eq. (S45) simplifies. Indeed, in this case, the corresponding adjacency matrix $\boldsymbol{A} = \{a_{ij}\}$ is irreducible [95]. Let us recall that for irreducible matrices the Perron–Frobenius theorem holds [76, 77], according to which there exists a positive eigenvalue $\widehat{\mu}$ greater or equal to (in absolute value) all other eigenvalues. Such eigenvalue corresponds to a simple root of the characteristic equation and the corresponding eigenvector $\vec{u}$ has all positive entries too. The latter vector is a multiple of the Bonacich eigenvector centrality vector [78]. Widely used in network science, the Bonacich eigenvector centrality is a measure that recursively accounts for local and global properties of the network, relying on the notion that a node can be highly central either by having a high degree or by being connected to others that themselves are highly central [62]. Simple algebraic steps can prove that if μ is an eigenvalue for $\boldsymbol{A}$, then $\lambda = f(\mu)$ is an eigenvalue for $\boldsymbol{M}$. Moreover, if $\vec{u}$ is an eigenvector corresponding to the eigenvalue μ of $\boldsymbol{A}$, then $\vec{u}$ is also an eigenvector corresponding to the eigenvalue $\lambda = f(\mu)$ of $\boldsymbol{M}$. Therefore, if $\widehat{\mu}$ is the maximum eigenvalue of

A, then $\widehat{\lambda} = f(\widehat{\mu}) = \frac{\nu}{\rho} + \frac{\nu+1}{\rho}\widehat{\mu} > 0$ is the highest eigenvalue of M, and with the same positive eigenvector $\vec{u}$. Thus, for strongly connected graphs, the approximated solution given by Eq. (S45) becomes

$$D_i(t) \underset{t\to\infty}{\approx} u_i\, t^{\widehat{\lambda}}, \tag{S46}$$

meaning that all nodes have similar Heaps' laws, and the key difference is made by their eigenvector centrality. As we saw in the main text (and we will see here more in details), these differences, more pronounced in transient times, will contribute to determine the fastest explorers in the network. Moreover, we deduce that the approximation used in Eq. (S42) is valid provided that $\widehat{\lambda} = f(\widehat{\mu}) < 1$, that is $\rho > \nu + (\nu+1)\widehat{\mu}$, while for higher values of ρ the solution is bounded by the linear solution as seen for the individual urn in Eq. (S7), since in the original system in Eq. (S40) we have $d_t D_i(t) \leq 1$.

Non-strongly connected network—Let us now consider the most general case, that is a directed or undirected graph with any hypotheses of connectivity. Let us construct an algorithm to determine the pace of discovery of each node, which will help us better understand analytically why some nodes have higher paces of discovery. To do this, let us partition the graphs into its strongly connected components (SCCs), i.e. maximal strongly connected subgraphs of G, which can be found in linear computational time, for example with a DFS-based algorithm [68]. Let all the SCCs be indexed as $C_1, \ldots, C_p$, with $C_i \cap C_j = \varnothing\ \forall i \neq j$.

Without loss of generality, let us suppose that the graph G is weakly connected, because otherwise we can repeat the same reasoning for each weakly connected component. Let us also suppose that the number of SCCs is $p > 1$, because otherwise the graph would be strongly connected, which we already discussed in the previous paragraph. Since G is weakly connected, for each SCC C_q there must exist another component C_l, with $l \neq q$, such that there are some links from C_q to C_l or viceversa. However, there cannot be links in both directions (from C_q to C_l and viceversa), because otherwise they would be a unique SCC. It is also easy to show that there is always a SCC without any outgoing links to other SCCs. Eventually permutating the indexes of the SCCs, let us call $C_1, \ldots, C_{p_1}$ all the components with no outer links. Then, for each $1 \leq q \leq p_1$, the respective system of differential equations for D_i, $i \in C_q$, does not depend on any outer variable D_j, $j \in C_l \neq C_q$. Therefore, we can consider C_q as an independent strongly connected subgraph of G, for which the reasoning in last paragraph holds. The solution for these SCCs is then:

$$D_i(t) \underset{t\to\infty}{\approx} \gamma_i^{(q)}\, t^{\widehat{\lambda}^{(q)}}\ \forall i \in C_q, 1 \leq q \leq p_1, \tag{S47}$$

where $\widehat{\lambda}^{(q)}$ is the maximum eigenvalue of the adjacency matrix of subgraph C_q and $\gamma_i^{(q)}$ is a multiple of the eigenvector centralitiy for node i in C_q. Found all the Heaps' laws relative to the nodes in $C_1, \ldots, C_{p_1}$, it is possible to show that there exist SCCs $C_{p_1+1}, \ldots, C_{p_2}$ that have links only towards the previously studied SCCs $C_1, \ldots, C_{p_1}$, with $p_2 > p_1$. Then, choosing C_q one of these other SCCs, let $\overline{\lambda}^{(q)}$ be the highest eigenvalue of the adjacency matrix of C_q. Let also $\tilde{\lambda}^{(q)} = \max_{l \leq p_1}(\gamma_{ql}\widehat{\lambda}^{(l)})$ be the maximum of the Heaps' exponents in Eq. (S47) of the SCCs reachable from C_q, where $\gamma_{qr} = 1$ if there is at least a link from C_q to C_l, $\gamma_{qr} = 0$ otherwise. As we will see further in this section, the Heaps' solutions for the nodes in these SCCs is:

$$D_i(t) \underset{t\to\infty}{\approx} \begin{cases} \gamma_i^{(q)}\, t^{\overline{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} > \tilde{\lambda}^{(q)} \\ \gamma_i^{(q)} \ln(t)\, t^{\tilde{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} = \tilde{\lambda}^{(q)} \quad \forall i \in C_q,\ p_1 + 1 \leq q \leq p_2, \\ \gamma_i^{(q)}\, t^{\tilde{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} < \tilde{\lambda}^{(q)} \end{cases} \tag{S48}$$

meaning that the Heaps' exponent $\widehat{\lambda}^{(q)}$ for node i in C_q, $p_1 + 1 \leq q \leq p_2$, is

$$\widehat{\lambda}^{(q)} = \max(\overline{\lambda}^{(q)},\ \tilde{\lambda}^{(q)}), \tag{S49}$$

that is the maximum of the highest eigenvalue $\overline{\lambda}^{(q)}$ of M relative to C_q and the highest $\tilde{\lambda}^{(q)}$ of the Heap's exponents $\widehat{\lambda}^{(l)}$ for $1 \leq l \leq p_1$. Moreover, if $\overline{\lambda}^{(q)} = \tilde{\lambda}^{(q)}$, a factor $\ln(t)$ appears in the solution. The same procedure can be repeated for all other successive SCCs C_q, keeping in mind that now a higher power $\ln^{\widehat{p}(q)}(t)$ of $\log(t)$ can appear.

In this algorithmic process, let us now consider a generic SCC, say C_q, and let us suppose we have solved inductively all the equations for the Heaps' law of the nodes in the already examined SCCs, that is $C_1, \ldots, C_{q-1}$. Let us recall that we arranged the indexes in such a way that the only outgoing links from C_q are pointed to nodes in previous SCCs, i.e. in some of the SCCs $C_1, \ldots, C_{q-1}$. For this reason, in order to solve the asymptotic differential equations responsible for the Heaps' law of the nodes in C_q, we can consider only the equations relative to the nodes in C_q in Eq. (S43), since the previous SCCs have been already solved and the following variables do not appear in these equations. We hence have to solve the following approximated equations:

$$\frac{dD_i(t)}{dt} \approx \frac{1}{t}\left(\frac{\nu}{\rho}D_i(t) + \frac{\nu+1}{\rho}\sum_{j\in C_q} a_{ij}D_j(t) + \frac{\nu+1}{\rho}\sum_{j\notin C_q} a_{ij}D_j(t)\right), \quad i \in C_q, \tag{S50}$$

where we have isolated the contributions coming from nodes outside C_q, which we have suppose to be known. Considering the general asymptotic solution for each individual Heaps' law derived for a strongly connected graph in Eq. (S45), for each $i \in C_q$ for large t we can write explicitly the functions $D_j(t)$, $j \notin C_q$, which lets us write:

$$\frac{\nu+1}{\rho} \sum_{j \notin C_q} a_{ij} D_j(t) \approx \frac{\nu+1}{\rho} \sum_{j \notin C_q} a_{ij} u_j \ln^{\widehat{p}_j}(t) \, t^{\widehat{\lambda}_j} \underset{t \to \infty}{\approx} \tilde{u}_i \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}} = f_i(t), \tag{S51}$$

where we have used the fact that $\eta_i \ln^{\tilde{p}_i}(t) \, t^{\tilde{\lambda}_i}$ is the leading term of the expression $\sum_{j \notin C_q} a_{ij} u_j \ln^{\widehat{p}_j}(t) \, t^{\widehat{\lambda}_j}$ and that we are working for large t. Then, using Eq. (S51) and calling $\vec{D}^{(q)}$ and $\boldsymbol{A}^{(q)}$ the sub-vector of $\vec{D}$ and sub-matrix of $\boldsymbol{M}$ relative to C_q, we can rewrite Eq. (S50) in a compact form as

$$\frac{d\vec{D}^{(q)}(t)}{dt} \approx \frac{\boldsymbol{M}^{(q)} \vec{D}^{(q)}(t)}{t} + \frac{\vec{f}^{(q)}(t)}{t}. \tag{S52}$$

The associated homogeneous system corresponds to the considering the sub-graph C_q without all the external links. For this system we get the same solution derived for a strongly connected graph in Eq. (S46), which is

$$\overrightarrow{\overline{D}}^{(q)}(t) \underset{t \to \infty}{\approx} \vec{\overline{u}}^{(q)} \, t^{\overline{\lambda}^{(q)}}, \tag{S53}$$

where $\overline{\lambda}^{(q)}$ is the highest eigenvalue of $\boldsymbol{M}^{(q)}$ (positive and simple for the Perron-Frobenious theorem), and $\vec{u}^{(q)}$ is a multiple of the eigenvector centrality. Let us search a solution for Eq. (S52) of the form $\vec{D}^{(q)}(t) = \vec{u}^{(q)}(t) \circ \overrightarrow{\overline{D}}^{(q)}(t)$, where $\circ$ is the Hadamard (element-wise) product, that plugged in Eq. (S52) gives:

$$\frac{d\vec{u}^{(q)}(t)}{dt} \circ \overrightarrow{\overline{D}}^{(q)}(t) + \cancel{\vec{u}^{(q)}(t) \circ \frac{d\left[\overrightarrow{\overline{D}}^{(q)}(t)\right]}{dt}} \approx \cancel{\vec{u}^{(q)}(t) \circ \frac{\boldsymbol{M}^{(q)} \overrightarrow{\overline{D}}^{(q)}(t)}{t}} + \frac{\vec{f}^{(q)}(t)}{t}, \tag{S54}$$

where the cancellation is due to the general solution in Eq. (S53) of the associated homogeneous system. Therefore, recalling Eq. (S51) and Eq. (S53) we have:

$$\frac{d\vec{u}^{(q)}(t)}{dt} \approx \vec{\tilde{u}}^{(q)} \circ \left(\vec{\overline{u}}^{(q)}\right)^{-1} \frac{\ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}}}{t^{\overline{\lambda}+1}} = \vec{\gamma} \frac{\ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}}}{t^{\overline{\lambda}+1}}, \tag{S55}$$

or equivalently, considering the i-th components:

$$\frac{du_i(t)}{dt} \approx \tilde{u}_i^{(q)} \left[\left(\vec{\overline{u}}^{(q)}\right)^{-1}\right]_i \frac{\ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}}}{t^{\overline{\lambda}+1}} = \gamma_i \frac{\ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}}}{t^{\overline{\lambda}+1}}, \tag{S56}$$

where we have defined $\vec{\gamma} = \vec{\tilde{u}}^{(q)} \circ \left(\vec{\overline{u}}^{(q)}\right)^{-1}$ and $\gamma_i = \tilde{u}_i^{(q)} \left[\left(\vec{\overline{u}}^{(q)}\right)^{-1}\right]_i$ its i-th component. Let us hence distinguish three cases.

1. If $\overline{\lambda}^{(q)} > \tilde{\lambda}^{(q)}$, then we have:

$$u_i(t) \approx \frac{\gamma_i}{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}} \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}} + u_i \underset{t \to \infty}{\approx} u_i, \tag{S57}$$

which gives the solution:

$$D_i(t) \underset{t \to \infty}{\approx} u_i t^{\overline{\lambda}^{(q)}}. \tag{S58}$$

2. Similarly, for $\overline{\lambda}^{(q)} = \tilde{\lambda}^{(q)}$ we have:

$$u_i(t) \approx \frac{\gamma_i}{\tilde{p}^{(q)} + 1} \ln^{\tilde{p}^{(q)}+1}(t) + u_i \underset{t \to \infty}{\approx} \frac{\gamma_i}{\tilde{p}^{(q)} + 1} \ln^{\tilde{p}^{(q)}+1}(t), \tag{S59}$$

which gives:

$$D_i(t) \underset{t \to \infty}{\approx} u_i \ln^{\tilde{p}^{(q)}+1}(t) \, t^{\tilde{\lambda}^{(q)}}. \tag{S60}$$

3. Finally, if $\overline{\lambda}^{(q)} < \tilde{\lambda}^{(q)}$ we have:

$$u_i(t) \approx \frac{\gamma_i}{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}} \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}} + d_1 \underset{t \to \infty}{\approx} \frac{\gamma_i}{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}} \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)} - \overline{\lambda}^{(q)}}, \tag{S61}$$

hence the solution:

$$D_i(t) \underset{t \to \infty}{\approx} a_i \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}}. \tag{S62}$$

To sum up, we have the following solutions:

$$D_i(t) \underset{t \to \infty}{\approx} \begin{cases} u_i t^{\overline{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} > \tilde{\lambda}^{(q)} \\ u_i \ln^{\tilde{p}^{(q)}+1}(t) \, t^{\tilde{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} = \tilde{\lambda}^{(q)} \quad \forall i \in C_q, q > p_1, \\ u_i \ln^{\tilde{p}^{(q)}}(t) \, t^{\tilde{\lambda}^{(q)}} & \text{if } \overline{\lambda}^{(q)} < \tilde{\lambda}^{(q)} \end{cases} \tag{S63}$$

Comparing this solution with the general one we gave in Eq. (S45), we have (a) $\hat{\lambda}(i) = \overline{\lambda}^{(q)}$ and $\hat{p}(i) = 0$ if $\overline{\lambda} > \tilde{\lambda}$, (b) $\hat{\lambda}(i) = \tilde{\lambda}^{(q)}$ and $\hat{p}(i) = \tilde{p}^{(q)} + 1$ if $\overline{\lambda} = \tilde{\lambda}$, and (c) $\hat{\lambda}(i) = \tilde{\lambda}^{(q)}$ and $\hat{p}(i) = \tilde{p}^{(q)}$ if $\overline{\lambda} < \tilde{\lambda}$.

In conclusion, when dealing with a network with multiple strength connected components, we solve the equations for the components that are independent from the others. Then we consider the SCCs that have links only to previous SCCs, applying the method just described. This is repeated until every SCC is studied, thus solving the whole system and describing the pace of discovery of each node of the entire network analytically, obtaining solutions of the type in Eq. (S45). In the next section this algorithmic method is applied to simple networks with $N = 4$ nodes, as we have already implicitly done above for a two nodes network and for chains.

G.   Application to the five graphs in Fig. 3

As an application of the analytical results of the previous sections, we study here the very same five networks reported in Fig. 3 of the main text. In particular, we will be able to provide an explicit expression for the growth of the number of novelties at each of the four nodes of the social network.

Graph a—Let us consider a network where nodes 2, 3, and 4 do not have any outgoing links, while node 1 has the links $1 \to 2$, $1 \to 3$, and $1 \to 4$ to all other nodes (see network representation in Table S1). Let us observe that the dynamics here is very similar to the case of a couple of urns with the only link $1 \to 2$. Nodes 2, 3, and 4 can be considered as three individual urns, for which the Heaps' law is the same to the classic one in Eq. (S9), that is:

$$D_2(t) \underset{t \to \infty}{\approx} D_3(t) \underset{t \to \infty}{\approx} D_4(t) \underset{t \to \infty}{\approx} (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}. \tag{S64}$$

As for node 1, the differential equation for the Heaps' law is approximated by:

$$\frac{dD_1(t)}{dt} \approx \frac{\nu D_1(t)}{\rho t} + \frac{(\nu + 1)\big(D_2(t) + D_3(t) + D_4(t)\big)}{\rho t} \approx \frac{\nu D_1(t)}{\rho t} + \frac{3(\nu + 1)(\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}}{\rho t}. \tag{S65}$$

The resolution of Eq. (S65) is the same as the one done for the couple of urns, with only a multiplicative factor 3. Therefore, the Heaps' solution for node 1 is:

$$D_1(t) \underset{t \to \infty}{\approx} 3\frac{\nu + 1}{\rho}(\rho - \nu)^{\frac{\nu}{\rho}} \ln(t) \, t^{\frac{\nu}{\rho}}, \tag{S66}$$

which means that node 1 has a higher pace of discovery than nodes 2, 3, and 4, but at asymptotic times they will show the same Heaps' exponent. Moreover, it is clear that in star-like networks adding more nodes does not increase significantly the pace of discovery.

Graph b—The next network we studied is a chain of 4 nodes, with links $1 \to 2$, $2 \to 3$, and $3 \to 4$. This network has already been studied in Sec. 1.3, and the solutions are:

$$D_i(t) \approx \frac{(\rho - \nu)^{\nu/\rho}}{(4 - i)!}\left(\frac{\nu + 1}{\rho} \ln(t)\right)^{4-i} t^{\nu/\rho}, \quad i = 1, 2, 3, 4. \tag{S67}$$

This analytical result shows us why node 1 has an higher pace of discovery than the other nodes, with lower Heaps' exponents for higher nodes. This is due to the presence of different powers of the logarithm. In the end, however, they all have the same asymptotic Heaps' exponent, meaning that the difference is visible only at finite times.

Graph c—Let us consider a network made by a directed cycle between nodes 2, 3 and 4, with links $2 \to 3$, $3 \to 4$, and $4 \to 2$, and another node 1 linked directly to node 2 ($1 \to 2$). In this case, we can distinguish two SCCs, the cycle and node 1. Since there is no link going out from the cycle, we start solving the Heaps' law equations related to it. As we have seen in Sec. 1.4, the solution is given by Eq. (S33) with N = 3, that is:

$$D_i(t) \underset{t\to\infty}{\approx} (\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{2\nu+1}{\rho}}, \quad i = 2, 3, 4. \tag{S68}$$

Now let us consider the remaining SCC, namely node 1. Its equation is the same as Eq. (S13) for the two coupled urns case in Sec. 1.2, with the only difference that here the solution of $D_2(t)$ has a higher exponent. Then, if we search for a solution like $D_1(t) = \kappa(t)\overline{D}_1(t)$, with $\overline{D}_1(t) \approx (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}$ being the solution of the associated homogeneous equation, we get:

$$\frac{d\kappa(t)}{dt} = \frac{\nu+1}{\rho t} \frac{D_2(t)}{\overline{D}_1(t)} \approx \frac{\nu+1}{\rho t} \frac{(\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{2\nu+1}{\rho}}}{(\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}} = \frac{\nu+1}{\rho} \frac{(\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{\nu+1}{\rho} - 1}}{(\rho - \nu)^{\frac{\nu}{\rho}}}, \tag{S69}$$

whose solution is:

$$\kappa(t) \approx \frac{\nu+1}{\rho} \frac{(\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{\nu+1}{\rho}}}{(\rho - \nu)^{\frac{\nu}{\rho}}}, \tag{S70}$$

which gives the asymptotic solution:

$$D_1(t) \approx \frac{\nu+1}{\rho} (\rho - 2\nu - 1)^{\frac{2\nu+1}{\rho}} t^{\frac{2\nu+1}{\rho}}. \tag{S71}$$

We could have obtained the same result using the algorithm developed in the last section. In this case, node 1 gets the same dynamics of the nodes in the cycle, with just a scaling factor $(\nu + 1)/\rho$, since the maximum eigenvalue of its SCC (node 1 itself) is lower than the maximum eigenvalue of the SCCs he is linked to (the cycle).

Graph d—In this case we consider the same network as the last graph we just analyzed swapping the direction of the link $4 \to 2$. Therefore, the cycle is broken (see network representation in Table S1), and as we are about to see, the dynamics is much more similar to a chain. We could give a detailed solution as done for the chain; instead, we are going to use directly the algorithm we developed to assess all the exponents in the Heaps' laws for every node. Let us start from node 4, which has no outgoing links. This node is hence an individual urn, with the usual solution:

$$D_4(t) \underset{t\to\infty}{\approx} (\rho - \nu)^{\frac{\nu}{\rho}} t^{\frac{\nu}{\rho}}. \tag{S72}$$

Let us move on to the SCC with outgoing links only towards previously studied SCCs, that is the SCC composed by node 3. If this SCC had no outgoing links, then it would be an isolated urn, therefore with the same exponent of the other SCC studied (node 4), meaning that the actual solution for node 3 has that exponent and a logarithmic factor. Indeed, the dynamics of node 3 is the same derived for the couple of urns in Sec. 1.2, which is:

$$D_3(t) \underset{t\to\infty}{\approx} \frac{\nu+1}{\rho} (\rho - \nu)^{\frac{\nu}{\rho}} \ln(t) \, t^{\frac{\nu}{\rho}}. \tag{S73}$$

Proceeding with node 2, we compare its exponent if it was isolated to the maximum of the exponents of node 3 and 4, which are all the same. Moreover, since node 3 has a higher power in the logarithm than node 4, in the asymptotic solution, we can disregard the presence of the link $4 \to 2$. Thus, the solution for node 2 has another logarithmic factor and another constant multiplicative factor than those of node 3, that is we have the solution:

$$D_2(t) \underset{t\to\infty}{\approx} \left(\frac{\nu+1}{\rho}\right)^2 (\rho - \nu)^{\frac{\nu}{\rho}} \ln^2(t) \, t^{\frac{\nu}{\rho}}. \tag{S74}$$

To complete, similarly we obtain the solution for node 1, i.e.:

$$D_1(t) \underset{t\to\infty}{\approx} \left(\frac{\nu+1}{\rho}\right)^3 (\rho - \nu)^{\frac{\nu}{\rho}} \ln^3(t) \, t^{\frac{\nu}{\rho}}. \tag{S75}$$

Network	(a)	(b)	(c)	(d)	(e)
$D_1(t) \underset{t\to\infty}{\approx}$	$u_1 \ln(t)\, t^{\frac{\nu}{\rho}}$	$u_1 \ln^3(t)\, t^{\frac{\nu}{\rho}}$	$u_1\, t^{\frac{2\nu+1}{\rho}}$	$u_1 \ln^3(t)\, t^{\frac{\nu}{\rho}}$	$u_1\, t^{\frac{\nu}{\rho}}$
$D_2(t) \underset{t\to\infty}{\approx}$	$u_2\, t^{\frac{\nu}{\rho}}$	$u_2 \ln^2(t)\, t^{\frac{\nu}{\rho}}$	$u_2\, t^{\frac{2\nu+1}{\rho}}$	$u_2 \ln^2(t)\, t^{\frac{\nu}{\rho}}$	$u_2\, t^{\frac{2\nu+1}{\rho}}$
$D_3(t) \underset{t\to\infty}{\approx}$	$u_3\, t^{\frac{\nu}{\rho}}$	$u_3 \ln(t)\, t^{\frac{\nu}{\rho}}$	$u_3\, t^{\frac{2\nu+1}{\rho}}$	$u_3 \ln(t)\, t^{\frac{\nu}{\rho}}$	$u_3\, t^{\frac{2\nu+1}{\rho}}$
$D_4(t) \underset{t\to\infty}{\approx}$	$u_4\, t^{\frac{\nu}{\rho}}$	$u_4\, t^{\frac{\nu}{\rho}}$	$u_4\, t^{\frac{2\nu+1}{\rho}}$	$u_4\, t^{\frac{\nu}{\rho}}$	$u_4\, t^{\frac{2\nu+1}{\rho}}$

TABLE S1. Summary of the asymptotic Heaps' laws derived analytically for the 4 nodes composing the five networks reported in Fig. 3 of the main text (here displayed at the top). The coefficients u_i have not been reported to focus on the exponents of the power laws and the logarithms, when present.

We can hence see that the solutions are equal to those of the chain in Sec. 1.7.b, and there are only some slight differences at finite times due to the presence of another link, but not significantly.

Graph e—The last case to be examined is again similar to Graph c, but this time we swap the direction of the link between nodes 1 and 2 (see network representation in Table S1). Here the order with which we study the SCCs is inverted, because now only node 1 has no outer links. Therefore, the Heaps' law for node 1 is the classic individual one in Eq. (S7). Then we have to solve the equations for the cycle, which in this case are:

$$
\begin{cases}
\dfrac{dD_2(t)}{dt} \approx \dfrac{\nu D_2(t)}{\rho t} + \dfrac{(\nu+1)D_3(t)}{\rho t} + \dfrac{(\nu+1)D_1(t)}{\rho t} \\[2mm]
\dfrac{dD_3(t)}{dt} \approx \dfrac{\nu D_3(t)}{\rho t} + \dfrac{(\nu+1)D_4(t)}{\rho t} \\[2mm]
\dfrac{dD_4(t)}{dt} \approx \dfrac{\nu D_4(t)}{\rho t} + \dfrac{(\nu+1)D_2(t)}{\rho t}.
\end{cases}
\tag{S76}
$$

In this system, we can consider $D_1(t)$ known, working at large time-scales. Therefore, following the algorithm described in Sec. 1.6.2, we first solve this system without the external sources (i.e. node 1), in order to find the leading solution and then compare the exponents with the external sources ones. The solution of the associated homogeneous system is the same of a directed cycle as in Eq. (S33), i.e. a power-law function with exponent $2\nu + 1/\rho$. Now, we observe that the Heaps' exponent of the cycle is higher than the exponents of outer SCCs it is linked to, that is just node 1 with exponent ν/ρ. Then, the asymptotic solution for the nodes in the cycle corresponds to the solution of the cycle as if it had no outer links. Explicit solutions are given in Table S1.

II. NODE RANKING AND HEAPS'LAW

In this section, we study more in details the validity of the eigenvector centrality and α-centrality to rank the nodes in a social network according to their discovery dynamics. First, we describe the real-world data sets considered. Then, we test the persistence of the nodes ranking based on the fitted Heaps' exponents at different times. Finally, we explain why the eigenvector centrality and the α-centrality lead to the same ranking of the Heaps' exponents for strongly-connected and generic networks respectively. All simulations in this section are performed with model parameters: $\rho = 10$, $\nu = 1$, $M_0 = \nu + 1$.

A. Description of the data sets

We consider four data sets of real-world networks representing different types of social interactions: the Zachary Karate Club (ZKC) network [74], a network of follower relationships among Twitter users [80], a co-authorship network in Network Science [81], and a collaboration network between jazz musicians [82]. The network of Twitter from the original data set (Ref. [80]) has been reduced by performing a random walk sampling.

Some basic properties of the networks are summarized in Table S2, like the total number of nodes N, the total number of links E, the average degree $\langle k \rangle$, and the maximum eigenvalue $\widehat{\mu}$ of the related adjacency matrix. Moreover, we have shown some properties of connection of the networks. In particular, we distinguished weakly-connected components (CCs) and strongly-connected components (SCCs), because they play an important role in the dynamics under investigation. Therefore we showed the number of both CCs and SCCs, as well as the size of the respective largest one. As we can see, the networks we have chosen have all very different properties, either in size, average degree, and connection.

Data set	Label	Type	N	E	$\langle k \rangle$	$\widehat{\mu}$	Num. CCs	Num. SCCs	Size LCC	Size LSCC
ZKC	(a)	Undirected	34	78	4.6	6.7	1	1	34	34
Twitter	(b)	Directed	4968	26875	10.8	5.2	1	4164	4968	770
NetSci	(c)	Undirected	1589	2742	3.4	19.0	396	396	379	379
Jazz	(d)	Undirected	198	2742	27.7	40.0	1	1	198	198

TABLE S2. Statistics and properties of the four real-world networks considered (see Fig. 4 of the main text): number of nodes N, number of edges E, average node degree k, maximum eigenvalue $\widehat{\mu}$, number of (weakly) connected components (CCs), and number of strongly connected components (SCCs), size of the largest (weakly) connected component (LCC), and size of the largest strongly connected component (LSCC).

B. Rank persistence

In the main text, we have developed a networked model for the dynamics of discovery that introduces an heterogeneity in the paces of discovery, as it happens in real-world social networks. In the previous sections, we concentrated on finding an analytical asymptotic solution of the Heaps' laws. However, for most of applications we are interested in transient times. As can be seen in Fig. 3 of the main text, the paces of discoverys, represented by the fitted Heaps' exponents, change in time, depending on the network characteristics and the model parameters. Nonetheless, the ranking of the nodes based on these fitted exponents remains almost the same. To show this, we plot in Fig. S1 the scatter plot and the Spearman's rank correlation coefficient between the fitted Heaps' exponents $\beta(T)$ at $T = 10^4$ and $T = 10^8$, together with their distributions, for the four real-world networks presented in the last section. In all cases, we get a Spearman's correlation of 1.00, meaning that even though the distribution of fitted exponents change, the ranking is time-invariant and does not depend on the particular T at which Heaps' exponents are fitted. Let us observe that we used a set of parameters that in all cases invalidate the approximations used in the analytical study, i.e. $\rho < \nu + (\nu + 1)\widehat{\mu}$.

This is evident in the scatter plot of Fig. S1(b), where, apart from a set of nodes whose exponents span across the entire range, most of the nodes present a very low pace of discovery, with fitted exponents very close to 0. A similar thing can be seen in Fig. S1(d), for which we have the highest eigenvalue and hence the highest Heaps' exponents among the four networks (with all Heaps' exponents very close to 1). All this is a strong indication that the various paces of discovery have to depend on some structural characteristics of the networks.

In the following sections, we keep investigating the relations between Heaps' exponents and network measures. In particular, we focus on the eigenvector centrality and the α-centrality, respectively useful for strongly-connected graphs and generic graphs. More insights on these centrality measures will be provided, both from a numerical and an analytical point of view.

C. Heaps' exponents and the eigenvector centrality

In the main text, we have shown that in strongly connected graphs each urn has the same asymptotic Heaps' exponent, and the driving factor for each node is the associated asymptotic coefficient. As we saw when we derived the asymptotic expression of the Heaps' law for strongly connected graphs in Eq. (S45), the Heaps' exponent corresponds to the maximum eigenvalue $\widehat{\lambda}$ of the matrix $\boldsymbol{M} = \frac{\nu}{\rho}\boldsymbol{I} + \frac{\nu+1}{\rho}\boldsymbol{A}$, where $\boldsymbol{A}$ is the adjacency matrix. In particular, because of the Perron-Frobenius theorem [76, 77], we know that $\widehat{\lambda}$ is positive and simple, and the related eigenvector $\vec{u}$ has all positive entries. We also derived that the coefficients of the Heaps' laws are all multiples of this eigenvector. A lot of importance has been given in the past to this vector, from which we can derive the eigenvector centrality, also known as the Bonacich centrality [78]. As a definition, the eigenvector centrality $c_i^{(E)}$ of node i is the i-th coefficient of the normalized solution of the equation:

$$\boldsymbol{M}\vec{c}^{(E)} = \widehat{\lambda}\,\vec{c}^{(E)}, \tag{S77}$$

where $\widehat{\lambda}$ is the highest positive eigenvalue [76]. This centrality measure accounts for both local and global properties of the network, as it is not just dependent on the degree of the node, but also on the positioning of each node in the network [69].
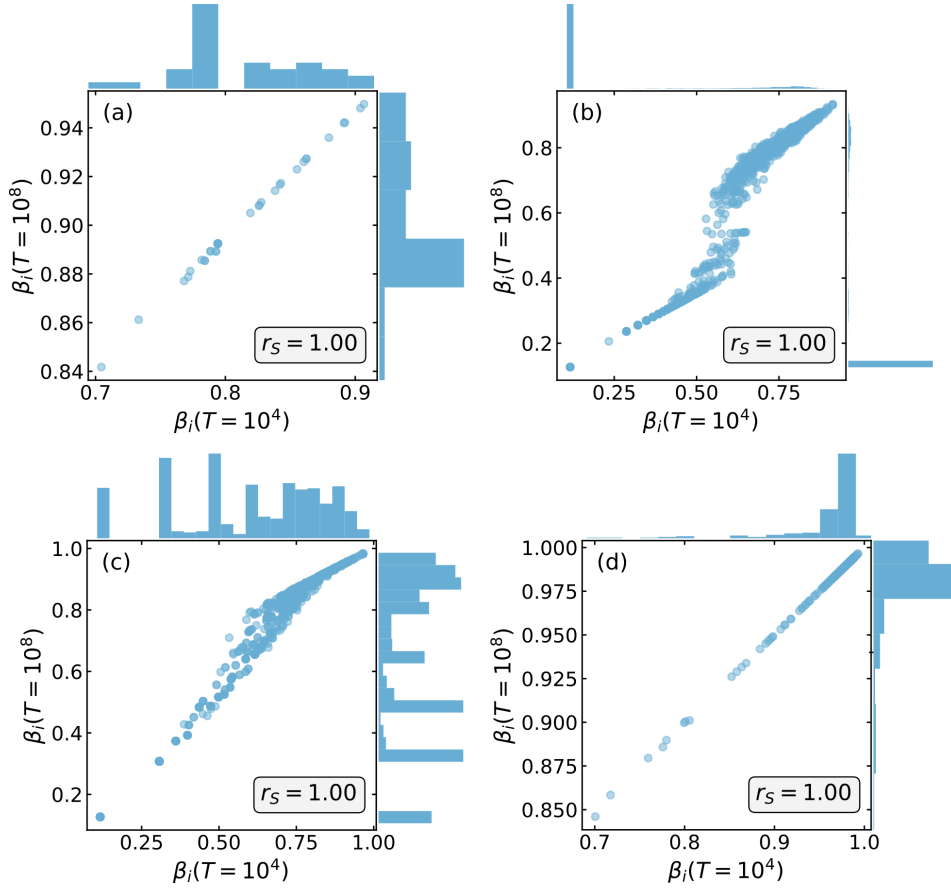
FIG. S1. Scatter plot and Spearman's rank correlation coefficient r_S between fitted Heaps' exponents $\beta_i(T)$ at $T = 10^4$ and $T = 10^8$ associated to the $i = 1, \ldots, N$ nodes off the four empirical networks considered: (a) the Zachary Karate Club network [74], (b) a network of follower relationships of Twitter [80], (c) a co-authorship network in network science [81] and (d) a collaboration network between jazz musicians [82]. The parameters of the model are $\rho = 10$, $\nu = 1$, $M_0 = \nu + 1$.

Our analytical investigation showed us that for strongly connected components we expect the same asymptotic Heaps' exponents. However, the same analysis showed us that the coefficients depend on the eigenvector centrality. This factor plays a role in the transient times, when we are far from the asymptotic regime, and it is thus especially important for real-world systems.

To complement the results presented in the main text, we now test numerically the correlation between the eigenvector centralities and the measured Heaps' exponents at transient times for the Zachary Karate Club network. Figure S2(a) shows the scatter plot and the Spearman's rank correlation of the eigenvector centralities and the fitted Heaps' exponents at time $T = 10^4$ for the (largest strongly connected component of) ZKC network and in Fig. S2(b) its visualization with color-coded nodes (cfr Fig. 2 of the main text). The resulting Spearman's rank correlation higher than 0.98 persists changing the parameters in the simulations, even for sets of parameters in contrast with the approximations used in the analytical study, i.e. $\rho < \nu + (\nu + 1)\widehat{\mu}$. We can hence conclude that the eigenvector centrality is an optimal proxy for the distribution of Heaps' exponents in strongly connected social networks, and it can be used to give a faithful ranking of the individual expected paces of discovery.

D. Heaps' exponents and the α-centrality

In this section we focus on generic directed graphs and the usage of the α-centrality as a proxy for the ranking of the nodes based on their pace of discovery in these more general cases. The α-centrality, widely used in network analysis [70–72], has been first introduced in Ref. [79] to extend the eigenvector centrality to asymmetric graphs. The underlying idea is to tune the influence of the adjacency matrix structure with a parameter α to add exogenous sources to the centrality [62, 79]. Formally, it is defined as the vector $\vec{u}$ such that

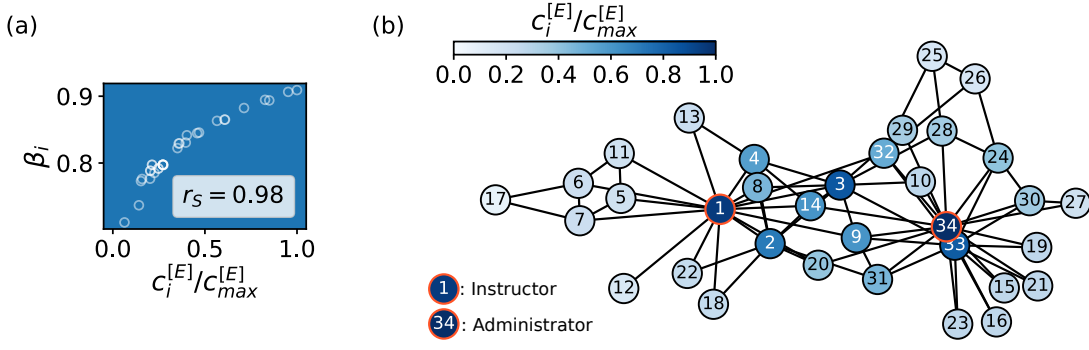$$\vec{c}^{(\alpha)} = \alpha \boldsymbol{A} \vec{c}^{(\alpha)} + \vec{e}, \tag{S78}$$

FIG. S2. Dynamics of the interacting urns on the Zachary Karate Club network [74]. (a) Scatter plot and Spearman's rank correlation coefficients r_S between fitted Heaps' exponents $\beta_i(T = 10^4)$ and normalized eigenvector centrality $c_i^{[E]}/c_{\max}^{[E]}$ associated to the $i = 1, \ldots, N$ nodes of the network. (b) Nodes are colored according to the resulting normalized eigenvector centrality.

where $\vec{e}$ is an N-dimensional vector of ones. The matricial form of Eq. (S78) reads:

$$\vec{c}^{(\alpha)} = (\boldsymbol{I} - \alpha\boldsymbol{A})^{-1}\vec{e} = \left(\sum_{k=0}^{\infty} \alpha^k \boldsymbol{A}^k\right)\vec{e}, \tag{S79}$$

where $\boldsymbol{I}$ is the N-dimensional identity matrix. It has also been shown that this centrality is equivalent to Katz-centrality [73] given by

$$\vec{c}^{(K)} = \left(\sum_{k=1}^{\infty} a^k \boldsymbol{A}^k\right)\vec{e}, \tag{S80}$$

with a being an attenuation factor. In fact, it has been shown that the equality $\vec{c}^{(K)} = -\vec{e} + \vec{c}^{(\alpha)}$ holds, i.e. these two centralities differ only by a constant [79]. From Eq. (S78) and (S79), it is clear that the α-centrality can be both a local and global measure. In fact, for $\alpha \to 0^+$, the relative importance of the structure given by the adjacency matrix $\boldsymbol{A}$ decreases, in favor of the exogenous factor given by $\vec{e}$. With higher values of α, instead, the role of the exogenous part is damped.

For an undirected graph, the α-centrality becomes proportional to the eigenvector centrality when $\alpha \to (1/\widehat{\mu})^-$, where μ is the highest positive eigenvalue of the adjacency matrix. In fact, in this case all eigenvalues are real and the eigenvectors are orthogonal. Following Ref. [79], let $\{\mu_\ell\}$ and $\{\vec{u}_\ell\}$ be the (eventually multiple) eigenvalues and eigenvectors of the adjacency matrix $\boldsymbol{A}$, with $\widehat{\mu} = \mu_1 > \mu_\ell$ for $\ell \neq 1$. Then we can write $\boldsymbol{A} = \sum_{\ell=1}^{N} \mu_\ell \vec{u}_\ell \vec{u}_\ell^T$. Considering that $\boldsymbol{A}^k = \sum_{\ell=1}^{N} \mu_\ell^k \vec{u}_\ell \vec{u}_\ell^T$, from Eq. (S79) we have:

$$\vec{c}^{(\alpha)} = \left(\sum_{k=0}^{\infty} \alpha^k \sum_{\ell=1}^{N} \mu_\ell^k \vec{u}_\ell \vec{u}_\ell^T\right)\vec{e} = \left(\sum_{\ell=1}^{N} \left(\sum_{k=0}^{\infty} \alpha^k \mu_\ell^k\right) \vec{u}_\ell \vec{u}_\ell^T\right)\vec{e} = \sum_{\ell=1}^{N} \frac{1}{1 - \alpha\mu_\ell} \vec{u}_\ell \vec{u}_\ell^T \vec{e}. \tag{S81}$$

When $\alpha \to (1/\widehat{\mu})^-$, the factor relative to $\ell = 1$ in the last term of Eq. (S81) becomes the leading term, thanks to the Perron-Frobenius theorem, so that we can write:

$$\lim_{\alpha \to (1/\widehat{\mu})^-} (1 - \alpha_1)\vec{c}^{(\alpha)} = (\vec{u}_1^T \vec{e})\vec{u}_1 \propto \vec{u}_1 \propto \vec{c}^{(E)}, \tag{S82}$$

where we have noted with $\vec{c}^{(E)}$ the eigenvector centrality.

Let us now generalize the analytical steps above to understand why the α-centrality correlates with the fitted Heaps' exponents for generic graphs, as we showed numerically in the main text for real-world social networks. Let us suppose that the social network is a weakly-connected directed graph, since otherwise we can repeat the same argument for each weakly-connected component. As we have shown before, the asymptotic behavior of the Heaps' law for node i is of the type $u_i \ln^{\widehat{p}(i)}(t)t^{\widehat{\lambda}(i)}$. We have shown also that the values of $\widehat{p}(i)$ and $\widehat{\lambda}(i)$ for each strongly connected component can be determined algorithmically. Here we will show that not only the α-centrality can account for the coefficient u_i like the eigenvector centrality, but also for the different values of $\widehat{p}(i)$ and $\widehat{\lambda}(i)$. Let us first concentrate on what happens with multiple eigenvalues, for which the biggest difference is primarily given by $\widehat{p}(i)$. Therefore, let us suppose for now that all SCCs in the graph have the same Heaps' exponent

$\widehat{\lambda}(i) = \widehat{\lambda}$, but different values of $\widehat{p}(i)$, and that in the leading terms the maximum value assumed by $\widehat{p}(i)$ is $\widehat{p}_{\max} < N$. This is the case for example of an open chain (already studied above), where $\widehat{p}(i) = 0, 1, \ldots, N - 1$ for $i = N, N - 1, \ldots, 1$ respectively, and $\widehat{p}_{\max} = N - 1$. Notice that, in this particular case, the adjacency matrix has only one eigenvalue $\widehat{\mu}$, related to the Heaps' exponent $\widehat{\lambda}$ through the relationship $\widehat{\lambda} = f(\widehat{\mu})$, with $f(x) = \frac{\nu}{\rho} + \frac{\nu+1}{\rho}x$. Therefore, the Jordan canonical form of the adjacency matrix is:

$$A = PJP^{-1} = P \begin{bmatrix} \widehat{\mu} & 1 & 0 & \cdots & 0 \\ 0 & \widehat{\mu} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \widehat{\mu} & 1 \\ 0 & \cdots & 0 & 0 & \widehat{\mu} \end{bmatrix} P^{-1} = P(\widehat{\mu}J_0 + J_1)P^{-1}, \tag{S83}$$

where $P = \left[\, \vec{u}_1 \,\middle|\, \vec{u}_2 \,\middle|\, \ldots \,\middle|\, \vec{u}_N \,\right]$ has the generalised eigenvectors in each column, and J_j denotes the $N \times N$ matrix with ones only in the $(j+1)$-th upper diagonal and null everywhere else, with $J_0 = I$. It is possible to show that

$$A^k = P(\widehat{\mu}J_0 + J_1)^k P^{-1} = P \left(\sum_{j=0}^{\min(N-1,\,k)} \binom{k}{j} \widehat{\mu}^{k-j} J_j \right) P^{-1} \tag{S84}$$

Hence, from Eq (S79), similarly to what we have done in Eq. (S81), we have:

$$\begin{aligned} \vec{c}^{(\alpha)} &= \left(\sum_{k=0}^{\infty} \alpha^k P(\widehat{\mu}J_0 + J_1)^k P^{-1} \right) \vec{e} = \left(\sum_{k=0}^{\infty} \alpha^k P \left(\sum_{j=0}^{\min(N-1,\,k)} \binom{k}{j} \widehat{\mu}^{k-j} J_j \right) P^{-1} \right) \vec{e} = \\ &= \left(\sum_{j=0}^{N-1} \left(\sum_{k=j}^{\infty} \alpha^k \binom{k}{j} \widehat{\mu}^{k-j} \right) P J_j P^{-1} \right) \vec{e} = \left(\sum_{j=0}^{N-1} \left(\sum_{k=0}^{\infty} \alpha^{k+j} \binom{k+j}{j} \widehat{\mu}^{k} \right) P J_j P^{-1} \right) \vec{e} = \\ &= \left(\sum_{j=0}^{N-1} \alpha^j \left(\sum_{k=0}^{\infty} \binom{k+j}{j} \alpha^k \widehat{\mu}^k \right) \sum_{\ell=1}^{N-j} \vec{u}_\ell \vec{u}_{\ell+j}^T \right) \vec{e} = \sum_{j=0}^{N-1} \sum_{\ell=1}^{N-j} \frac{\alpha^j}{(1 - \alpha\widehat{\mu})^{j+1}} \vec{u}_\ell \vec{u}_{\ell+j}^T \vec{e} = \\ &= \sum_{\ell=1}^{N} \left(\sum_{j=0}^{N-\ell-1} \frac{\alpha^j}{(1 - \alpha\widehat{\mu})^{j+1}} \vec{u}_{\ell+j}^T \vec{e} \right) \vec{u}_\ell. \end{aligned} \tag{S85}$$

From the above, it is clear that the nodes ℓ for which $(\vec{u}_1)_\ell$ is positive have the greatest α-centrality when $\alpha \to 1/\widehat{\mu}$, since they they are associated to the highest power in the logarithm $\widehat{p}(i) = \widehat{p}_{\max}$. Among these, as with the eigenvector centrality, nodes with higher coefficients (corresponding to the eigenvector centralities in that SCC) have higher ranking. Then the nodes who have zeroes in $\vec{u}_1$ but positive entries in $\vec{u}_2$ are next in the ranking, and so on. This confirms the fact that, when comparing nodes with same asymptotic Heaps' exponent, those with higher discovery rates, i.e. those with higher powers in the logarithm factor, have the highest α-centrality.

A similar approach to the one we used to derive the algorithmic solution of the Heaps' law for a generic graph can be used to treat generic weakly-connected graphs. Let us divide the network into its SCCs. For each component C_q, we denote $\mu^{(q)}$ the maximum between the maximum eigenvalue the component would have if isolated and the maximum eigenvalue of the neighboring SCCs, following the same order used with the developed algorithm. In this setting, it is then possible to compute the α-centrality at $\alpha \to (1/\mu^{(q)})^-$, that might be different across SCCs. The final ranking is given by ordering the evaluated α-centralities starting from those with the highest $\mu^{(q)}$.

It is worth noticing that this method can be computationally not efficient, especially for big networks. For this reason, we have tested how reliable the α-centrality with the same value of α is when comparing it to the Heaps' exponents, regardless of the procedure above. In Fig. 4 of the main text we have investigated the relation between Heaps' exponents and α-centralities setting α to $0.85/\widehat{\mu}$. Here, we further investigate how the correlation changes with α. This is shown in Fig. S3, where we plot the Spearman's rank correlation coefficient between the paces of discovery $\beta_i(10^4)$ and the α-centralities $c_i^{[\alpha]}$ as a function of α for all the nodes $i = 1, \ldots, N$ composing the four considered real-world networks. Although panel (d) displays a decrease in the correlation when approaching $1/\widehat{\mu}$, however, setting $\alpha < 1/\widehat{\mu}$ leads to Spearman's rank correlation coefficients $r_S > 0.89$ in all four cases (cfr main text).
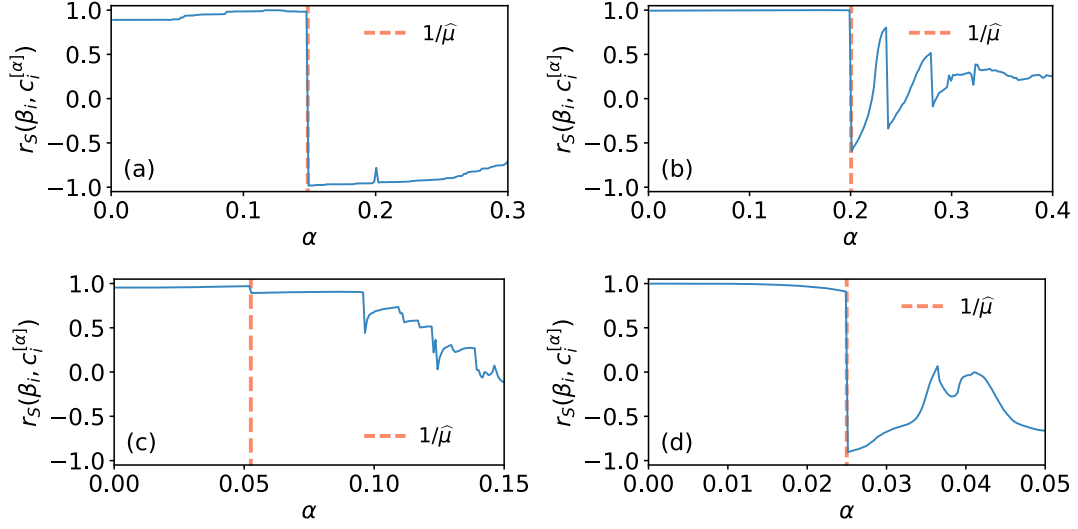
FIG. S3. Spearman's rank correlation r_S between paces of discovery $\beta_i(10^4)$ and α-centrality $c_i^{[\alpha]}$ as a function of α for nodes $i = 1, \ldots, N$ belonging to four different real-world networks: (a) the Zachary Karate Club network [74], (b) a network of follower relationships of Twitter [80], (c) a co-authorship network in network science [81] and (d) a collaboration network between jazz musicians [82]. Each dashed vertical line corresponds the value of $1/\widehat{\mu}$, with $\widehat{\mu}$ denoting the maximum eigenvalue of the corresponding adjacency matrix. The parameters of the model are $\rho = 10$, $\nu = 1$, $M_0 = \nu + 1$.