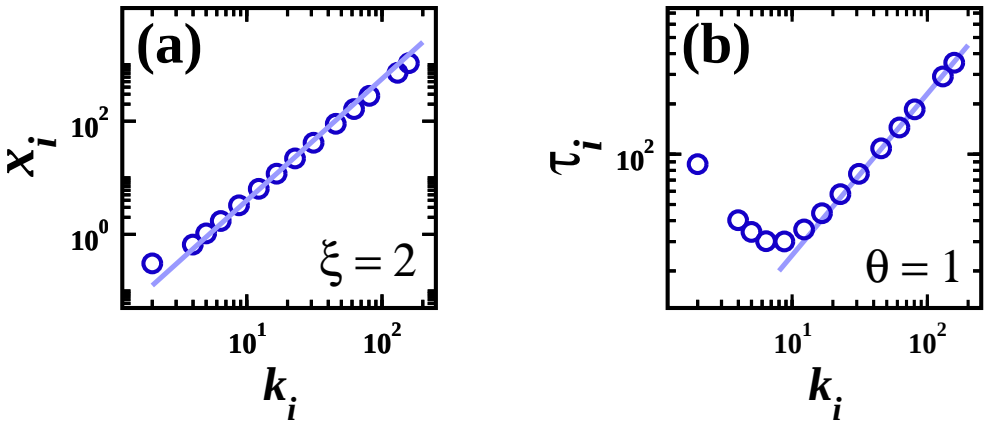
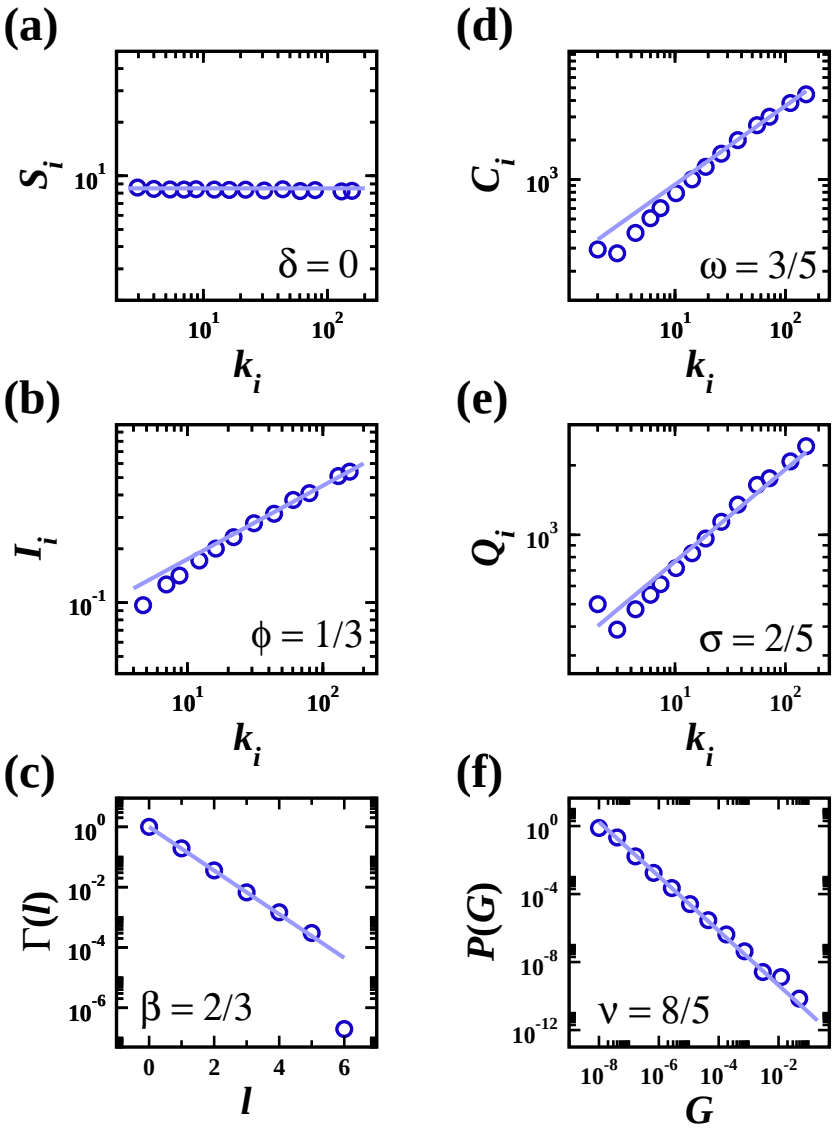
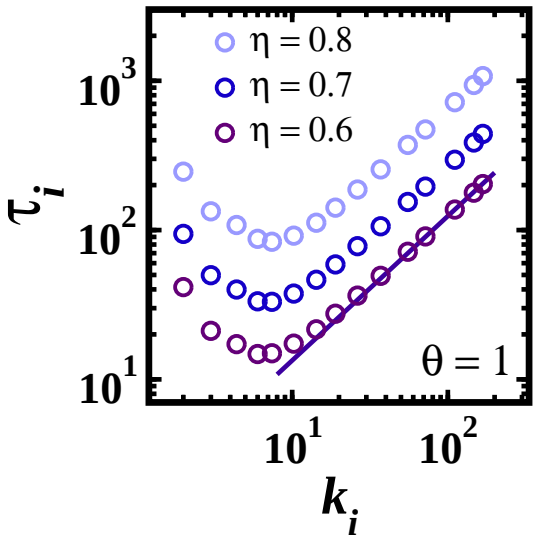


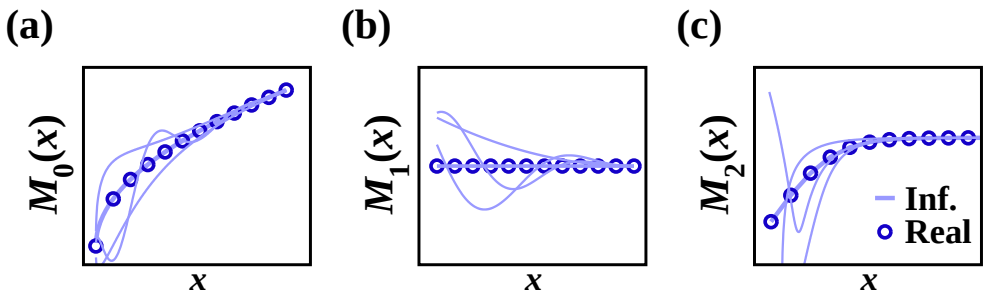
Supplementary Figure 1: **The transient observation for regulatory dynamics**. (a) The steady state x_i vs. degree k_i (circles) as obtained from Eq. (18) on a scale free network ($N = 6,000, \langle k \rangle = 8, P(k) \sim k^{-\gamma}$ with $\gamma = 3$, see Supplementary Note 4.1 - 2). The activities are shown to scale as $x_i \sim k_i^{\xi}$ where $\xi = 2$ (solid line), as predicted in (21). (b) The relaxation time scales as $\tau_i \sim k_i^{\theta}$ in the limit of large k_i, where $\theta = 1$ in agreement with the prediction of (45).



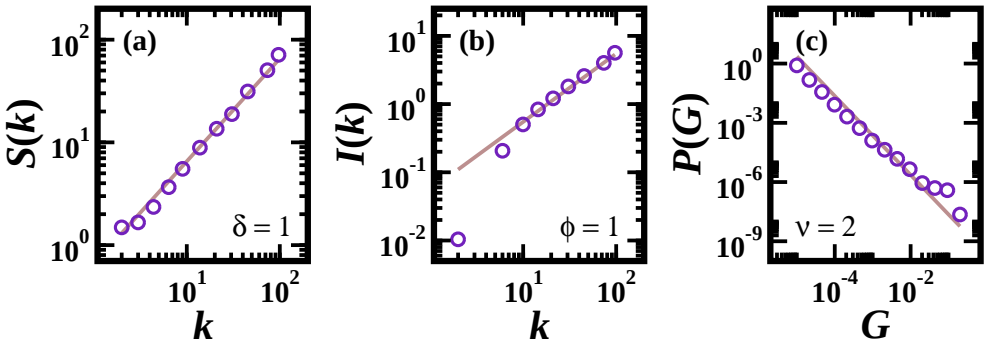
Supplementary Figure 2: **The asymptotic observation for regulatory dynamics**. (a) The stability S_i has $\delta = 0$ (51); (b) the impact I_i has $\varphi = 1/3$ (54); (c) the propagation function $\Gamma(l)$ decays as $\Gamma(l) = e^{-\beta\alpha l}$ with $\beta = 2/3$ (59) as predicted. From these three functions we derive additional observations: (d) the cascades C_i have $\omega = 3/5$ (63); (e) the incoming cascades Q_i have $\sigma = 2/5$ (72); (f) the response distribution $P(G)$ follows a power law with $\nu = 8/5$ (61). Simulations were carried out on a scale free network, $N = 6,000$, $\langle k \rangle = 8$ and $\gamma = 3$ (Supplementary Note 4.1 - 2).



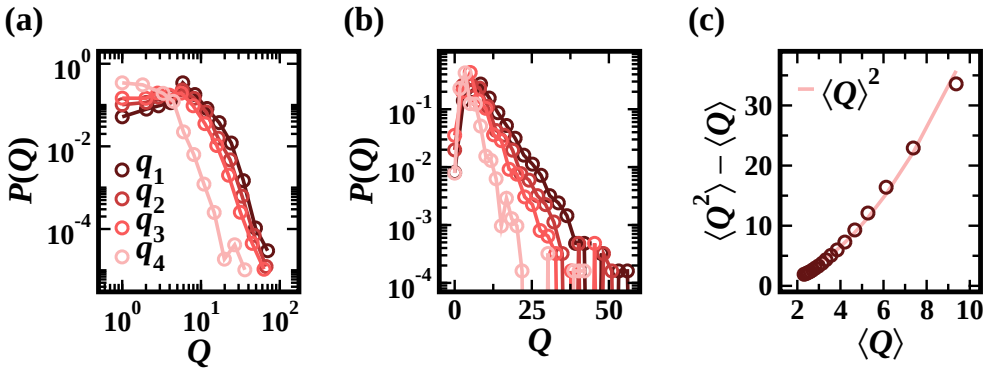
Supplementary Figure 3: **Measuring the relaxation time**. To observe $\tau(k)$ we measured the temporal response of all nodes to perturbations, taking the relaxation time of i to be the time it takes for it to reach a fraction η of its final perturbation $dx_i(t \to \infty)$ (31). We set $\eta = 0.6, 0.7, 0.8$, and observe that the scaling exponent θ is independent of this selection.



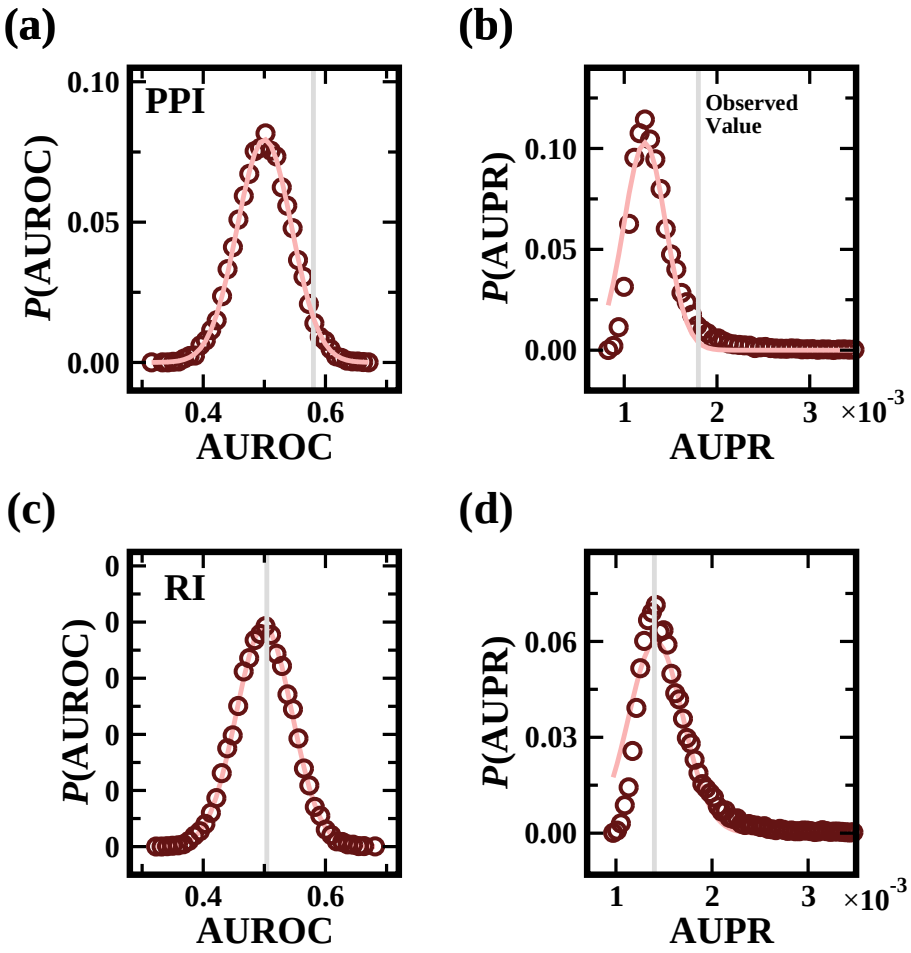
Supplementary Figure 4: **Inference of regulatory dynamics**. We used observations $\mathcal{T}$ (Supplementary Figure **1**) and $\mathcal{G}$ (Supplementary Figure **2**) to infer the dynamical functions $M_0(x)$, $M_1(x)$ and $M_2(x)$ of (18). (a) For $M_0(x)$ our inference predicts that asymptotically $M_0(x)|_{x\to\infty} \sim x^{1/2}$. This captures the true behavior of the function which is $M_0(x) = Bx^{1/2}$ (circles). As there is still a degree of freedom to add lower order terms our inference predicts a *class* of potential functions, defining the minimal subspace $\mathbb{M}(\mathcal{T},\mathcal{G})$, rather than a single $M_0(x)$. Several possible functions from $\mathbb{M}(\mathcal{T},\mathcal{G})$ are shown (solid lines), all of which behave asymptotically like the real $M_0(x)$, and are guaranteed to recover the observed exponents. (b) For $M_1(x)$ we predict $M_1(x)|_{x\to\infty} \sim 1$, capturing the asymptotic behavior of the actual function used (circles). (c) For $M_2(x)$ the inference provides the asymptotic behavior $M_2(x)|_{x\to\infty} \sim 1 - x^{1/3}$. Hence all potential functions in $\mathbb{M}(\mathcal{T},\mathcal{G})$ capture the defining feature of regulatory dynamics that regulation saturates in the limit of large x.



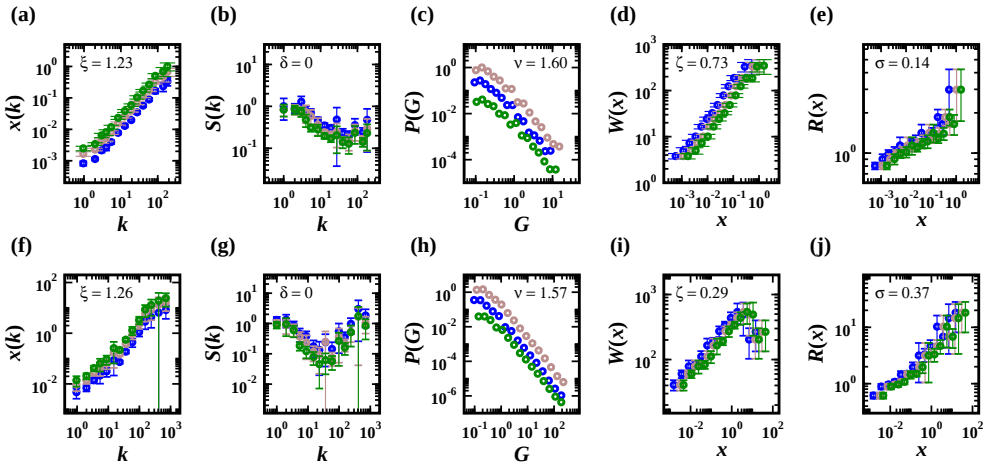
Supplementary Figure 5: **Reverse engineering Kuramoto dynamics**. To test our formalism against non-factorizable dynamics we simulated Eq. (115) on a scale free A_{ij} and measured the system's response to perturbations via (117). We find that (a) $S(k) \sim k^{\delta}$ where $\delta = 1$; (b) $I(k) \sim k^{\varphi}$ where $\varphi = 1$; (c) $P(G) \sim G^{-\nu}$ where $\nu = 2$.



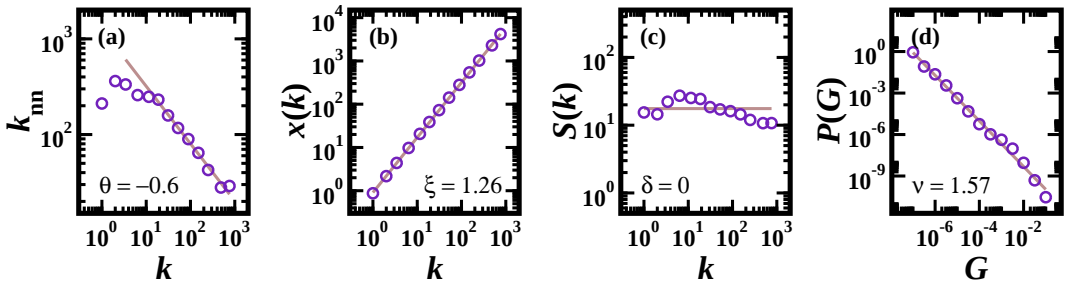
Supplementary Figure 6: **The incoming cascade size distribution of cellular perturbations**. (a) We measured $P(Q)$ from the genetic perturbation experiment[1], for different values of the threshold q (64), $q_1 = 0.1$, $q_2 = 0.25$, $q_3 = 0.5$ and $q_4 = 0.75$. The sharp drop for large Q indicates that $P(Q)$ is bounded. (b) Using semi-logarithmic axes we find that for large Q the distribution features an exponentially decaying tail. (c) The variance vs. the average of $P(Q)$ as obtained for different values of q. The data are well approximated by a quadratic function (solid line).



Supplementary Figure 7: **Evaluating the significance of AUROC and AUPR**. By measuring AUROC and AUPR on 25,000 randomly generated predictions we obtain the statistical significance of the observed evaluation curves. **Protein-protein interactions**: (a) The probability density function $P(\text{AUROC})$ (circles) follows a normal distribution (solid line). The observed AUROC for PPI (grey solid line) has a z-score of ~ 2, corresponding to a p-value of ~ 0.03. (b) $P(\text{AUPR})$ can be approximated by a normal distribution around the mean, however is features a tail which decays slower than that of a normal distribution for large AUPR. Hence to evaluate the significance of the observed AUPR (grey solid line) we measured the fraction of randomized predictions for which AUPR exceeded the observed value (namely the area under $P(\text{AUPR})$ to the right of the grey solid line), providing p-value ~ 0.07. **Regulatory interactions**: (c) - (d) For RI the measured AUROC and AUPR are statistically insignificant compared to randomized predictions (indicated by the location of the grey solid line around the median of the distribution).



Supplementary Figure 8: **The impact of Δt on the dynamical functions of human dynamics**. (a) - (e) To test the effect of the selection of Δt on the empirical observations extracted from UCIonline we constructed G_{ij} using $\Delta t = 3$ (green), 6 (brown) and 9 hours (blue). The measured exponents are shown to be independent of Δt. (f) - (j) The same for Email Epoch, using $\Delta t = 6$ (green), 9 (brown) and 12 hours (blue).



Supplementary Figure 9: **Reverse engineering in the presence of degree correlations**. (a) The average degree of a node's nearest neighbors vs. its own degree, k (logarithmically binned), as obtained from Email Epoch. The scaling relationship $k_{\rm nn}(k) \sim k^{-0.6}$ in the limit of large k, indicates the significant level of degree correlations exhibited by this network. (b) - (d) We numerically simulated the reverse engineered human dynamics (152) on Email Epoch and obtained $\mathcal{G}$ and $\mathcal{T}$ for this system (following the procedure described in Supplementary Note 4). We find that, in spite of the significant degree correlations, the numerically predicted ξ, δ and ν are in perfect agreement with those observed empirically for this system (Fig. 5, main text). Hence degree correlations have no significant effect on the predictive power of our formalism.

Supplementary Note 1. Transient Observation $\mathcal{T}$

We consider a system of N nodes, whose activities $x_i(t)$ $(i = 1, \ldots, N)$ are driven by the dynamic equation

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = M_0(x_i) + \sum_{j=1}^{N} A_{ij} M_1(x_i) M_2(x_j). \tag{1}$$

The first term on the r.h.s. describes the self-dynamics, *i.e.* a node's behavior in the absence of interactions; the second term sums over all of i's interacting partners j. The adjacency matrix A_{ij} represents the underlying network of interactions, which, in our analytic calculations, we take to be a *configuration model* network, namely a network with an arbitrary degree sequence, in which the nodes are linked at random[3,4]. In the configuration model, since the network is constructed from an arbitrary degree sequence, all forms of degree distributions $P(k)$ are supported, including bounded (*e.g.*, Poisson) or fat-tailed (*e.g.*, scale-free) distributions. However, the random process by which the links are drawn ensures that A_{ij} has little structural correlations, such as degree-degree correlations or clustering. Hence, while the nodes themselves could potentially exhibit a high level of heterogeneity in their degrees k_i, depending on $P(k)$, their surrounding, *e.g.*, the average degrees of nodes in their neighborhood, are on average the same, independent of k_i.

1.1 Steady state

As we are interested in the response of the system to perturbations around the steady state, we consider systems of the form (1) that exhibit at least one fully positive steady state x_i $(i = 1, \ldots, N)$. We obtained this steady state by setting the derivative on the l.h.s. to zero. To characterize the degree dependence of x_i, we denote by $\mathbf{E}(k)$ the group of all nodes with degree k and calculate the activity of the average node in that group, namely

$$x(k) = \frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} x_i, \tag{2}$$

where $|\mathbf{E}(k)|$ represents the number of nodes in $\mathbf{E}(k)$. Using (1) we write

$$\frac{\mathrm{d}x(k)}{\mathrm{d}t} = \frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} \Big[M_0(x_i) + \sum_{j=1}^{N} A_{ij} M_1(x_i) M_2(x_j) \Big], \tag{3}$$

which we approximate by

$$\frac{\mathrm{d}x(k)}{\mathrm{d}t} = M_0\big(x(k)\big) + M_1\big(x(k)\big) \frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} \sum_{j=1}^{N} A_{ij} M_2(x_j). \tag{4}$$

Equation (4) is exact in the limit where

$$\frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} M_q(x_i) \approx M_q \left(\frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} x_i \right), \tag{5}$$

($q = 0, 1$). The sum on the r.h.s. of (4) represents an average of $M_2(x_j)$ of nodes j that are nearest neighbors of nodes in $\mathbf{E}(k)$. However, the lack of structural correlations in the configuration model implies that this average is independent of k. Indeed, due to the random process by which links are drawn, nearest neighbors of $\mathbf{E}(k)$ are on average no different than nearest neighbors of any other group, $\mathbf{E}(k')$. Hence we can replace this sum by an average over nearest neighbors in general, namely

$$\frac{1}{|\mathbf{E}(k)|} \sum_{i \in \mathbf{E}(k)} \sum_{j=1}^{N} A_{ij} M_2(x_j) = k \left\langle M_2(x) \right\rangle_{\mathbf{E}(k)} = k \left\langle M_2(x) \right\rangle_{\odot}, \tag{6}$$

where

$$\langle X \rangle_{\odot} = \frac{1}{N} \sum_{i=1}^{N} \frac{1}{k_i} \sum A_{ij} X_j \tag{7}$$

represents an average over *all* nearest neighbor nodes. Using this notation we can write Eq. (4) for an average node of degree k as

$$\frac{\mathrm{d}x(k)}{\mathrm{d}t} = M_0\big(x(k)\big) + k M_1\big(x(k)\big) \left\langle M_2(x) \right\rangle_{\odot}. \tag{8}$$

At steady state we set the l.h.s. of (8) to zero, obtaining

$$R\big(x(k)\big) = \frac{1}{\langle M_2(x)\rangle_{\odot}\, k}, \tag{9}$$

where

$$R(x) = -\frac{M_1(x)}{M_0(x)}. \tag{10}$$

Hence using (9) we find

$$x(k) \sim R^{-1}(\lambda), \tag{11}$$

where $R^{-1}(x)$ is the inverse function of $R(x)$ and

$$\lambda = \frac{1}{\langle M_2(x)\rangle_{\odot}\, k} \sim k^{-1}. \tag{12}$$

For large k (small λ) we take only the leading terms of (11). Hence we express $R^{-1}(\lambda)$ as a Hahn series[5]

$$R^{-1}(\lambda) = \sum_{n=0}^{\infty} r_n \lambda^{\Pi_R(n)}. \tag{13}$$

The powers that participate in (13) are represented by Π_R, which is a well ordered set of real numbers. The nth power, $\Pi_R(n)$ is at placement n in the ordering, namely $\Pi_R(n-1) < \Pi_R(n) < \Pi_R(n+1)$; the leading power of (13) is thus $\Pi_R(0)$. In case $\Pi_R(0) \neq 0$ the behavior of (13), and hence (11), in the limit of small λ is governed by this power, namely $x(k) \sim \lambda^{\Pi_R(0)}$. However, when $\Pi_R(0) = 0$ the leading order of (13) is a constant $R^{-1}(\lambda) \sim r_0 + \ldots$, and the λ dependence of (13) and (11) is captured by the next, non-vanishing power $\Pi_R(1) > 0$. Hence $x(k)$ (11) can observe two classes of behavior:

Scaling: if in (13) the leading power, $\Pi_R(0)$, is non-vanishing, namely $R^{-1}(\lambda) \sim \lambda^{\Pi_R(0)}$ up to higher order terms in λ, we have from (11)

$$x(k) \sim k^{\xi}, \tag{14}$$

(to leading order) where

$$\xi = -\Pi_R(0). \tag{15}$$

Here, in the limit of large k, the steady state activity of a node scales with its degree, increasing (decreasing) with k for $\xi > 0$ ($\xi < 0$).

Saturation: the only other possibility is that $\Pi_R(0) = 0$, in which case $R^{-1}(\lambda) \sim r_0 + r_1 \lambda^{\Pi_R(1)}$, where $\Pi_R(1) > 0$. Here the activities saturate as

$$x(k) \sim x_0 - k^{\xi} \tag{16}$$

where

$$\xi = -\Pi_R(1). \tag{17}$$

Here, as $\Pi_R(1) > 0$, we have $\xi < 0$, hence $x(k)$ approaches a maximum activity of x_0 in the limit of large k. Note that the coefficients in (16), x_0 and -1, are chosen arbitrarily, as we are only focused on the scaling/saturating behavior of (14) and (16), and not on the specific coefficients of the expansion. Indeed choosing $x(k) \sim x_0 + k^{\xi}$ in (16) is equally justified mathematically. Our choice here reflects the common scenario, where, in practice, activities increase with k, until approaching an upper bound of $x \to x_0$.

Example - regulatory dynamics: To demonstrate the prediction of $x(k)$ (11) we refer to the regulatory model (Hill function)[6,7] introduced in the main text. Here (1) takes the form

$$\frac{\mathrm{d}x_i}{\mathrm{d}t} = -Bx_i^a + \sum_{j=1}^{N} A_{ij} \frac{x_j^h}{1 + x_j^h}, \tag{18}$$

where B is a constant, and we set $a = 1/2$ and $h = 1/3$. The dynamical functions are $M_0(x) = -Bx^a$, $M_1(x) = 1$ and $M_2(x) = x^h/(1 + x^h)$. Using Eq. (10) we write

$$R(x) = \frac{1}{B} x^{-a}, \tag{19}$$

allowing us to write (13) as

$$R^{-1}(\lambda) = B^{-\frac{1}{a}} \lambda^{-\frac{1}{a}}. \tag{20}$$

Equation (20) is a specific case of the Hahn expansion (13) in which $\Pi_R = \{-1/a\}$, comprising a single non vanishing power. Hence the leading power in the expansion is $\Pi_R(0) = -1/a \neq 0$, predicting that $x(k)$ follows (14), with $\xi = 1/a = 2$, namely

$$x(k) \sim k^2. \tag{21}$$

In Supplementary Figure **1**a we display $x(k)$ vs. k as obtained from regulatory dynamics (18), confirming the validity (21).

1.2 Temporal dynamics

We now turn to characterizing the system's response to perturbations. We induce a small permanent perturbation $\mathrm{d}x_m$ on the steady state activity of node m. The dynamical equation (1) remains the same for all nodes except for node m, which under the permanent perturbation no longer evolves with time, hence is follows

$$x_m(t) = x_m + \mathrm{d}x_m. \tag{22}$$

For the remaining nodes (1) provides

$$\frac{\mathrm{d}}{\mathrm{d}t}(x_i + \mathrm{d}x_i) = M_0(x_i + \mathrm{d}x_i) + \sum_{j=1}^{N} A_{ij} M_1(x_i + \mathrm{d}x_i) M_2(x_j + \mathrm{d}x_j). \tag{23}$$

Linearizing around the steady state we obtain

$$\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}x_i) &= \left(M_0'(x_i) + M_1'(x_i) \sum_{j=1}^{N} A_{ij} M_2(x_j) \right) \mathrm{d}x_i(t) \\ &+ M_1(x_i) \sum_{j=1}^{N} A_{ij} M_2'(x_j) \mathrm{d}x_j(t) + \mathcal{O}(\mathrm{d}x^2), \end{aligned} \tag{24}$$

where $M_q'(x)$ $(q = 0, 1, 2)$ represents the derivative $\mathrm{d}M_q/\mathrm{d}x$ with x taken at the steady state, namely $x = R^{-1}(\lambda)$ (11). Consider the two summation terms at the r.h.s. of (24), both of which sum over i's nearest neighbors. Using the same principles that lead to (6), we rely on the configuration model to write the first sum as

$$\sum_{j=1}^{N} A_{ij} M_2(x_j) = k_i \left\langle M_2(x) \right\rangle_{\odot}, \tag{25}$$

replacing the average over i's nearest neighbors by the general nearest neighbor average $\langle \cdot \rangle_{\odot}$ (7). The treatment of the second sum in (24) requires more caution, due to its time dependence: indeed, while the state of i's neighborhood is independent of i, it time dependent response $\mathrm{d}x_j(t)$ does depend on i's distance l_{im} from the source of the perturbation m. This is because the perturbation must propagate along paths of length l_{im} from m to the surrounding nodes of i. Hence we write

$$\sum_{j=1}^{N} A_{ij} M_2'(x_j) \mathrm{d}x_j(t) = k_i \left\langle M_2'(x) \mathrm{d}x(t) \right\rangle_{\odot, l_m}, \tag{26}$$

where

$$\langle X \rangle_{\odot, l_m} = \frac{1}{|K_m(l)|} \sum_{i \in K_m(l)} \frac{1}{k_i} \sum A_{ij} X_j \tag{27}$$

represents the nearest neighbor average among nodes at distance l from the source m. In (27) $K_m(l)$ denotes the group of all nodes at distance l from m and $|K_m(l)|$ is the number of nodes in that group. Using (25) and (26) we can now write (24) as

$$\frac{\mathrm{d}}{\mathrm{d}t}(\mathrm{d}x_i) = -\frac{1}{\tau_i} \mathrm{d}x_i + \mathcal{B}_i(l, t) \tag{28}$$

where

$$\tau_i = -\frac{1}{M_0'(x_i) + k_i M_1'(x_i) \left\langle M_2(x) \right\rangle_{\odot}} \tag{29}$$

and

$$\mathcal{B}_i(l, t) = k_i M_1(x_i) \left\langle M_2'(x) \mathrm{d}x(t) \right\rangle_{\odot, l}. \tag{30}$$

Equation (28) is a non-homogeneous linear differential equation, describing the time dependent perturbation $\mathrm{d}x_i(t)$, of a node with degree k_i at distance l from a permanent perturbation $\mathrm{d}x_m$. Its solution takes the form

$$\mathrm{d}x_i(t) = C e^{-\frac{t}{\tau_i}} + e^{-\frac{t}{\tau_i}} \int_0^t \mathcal{B}_i(l, t') e^{\frac{t'}{\tau_i}} \mathrm{d}t', \tag{31}$$

where C is a constant set to zero to satisfy the initial condition $\mathrm{d}x_i(t = 0) = 0$. The relaxation of $\mathrm{d}x_i$ (31) to its final, perturbed, state is governed by two independent time scales: We denote the first by $\tau_{\odot}$, representing the relaxation time of $\mathcal{B}_i(l, t)$, which depends on the response time of i's nearest neighbors, and hence it is independent of i. The second time scale, τ_i (29), represents i's intrinsic relaxation time, which depends on k_i.

To analyze τ_i's dependence on k_i, or $\tau(k)$, we focus on each of the two terms in the denominator of (29) independently. First we write

$$M_0' = \left. \frac{\mathrm{d}M_0(x)}{\mathrm{d}x} \right|_{x = R^{-1}(\lambda)}, \tag{32}$$

where we used (11) to represent the steady state value $x(k)$. Using the definition of $R(x)$ (10) we continue writing

$$\begin{aligned} M_0' &= \left(-\frac{M_1'(x)}{R(x)} + \frac{M_1(x)}{R^2(x)}R'(x)\right)\Big|_{x=R^{-1}(\lambda)} \\ &= -\frac{M_1'\big(R^{-1}(\lambda)\big)}{\lambda} + \frac{M_1\big(R^{-1}(\lambda)\big)R'\big(R^{-1}(\lambda)\big)}{\lambda^2}, \end{aligned} \tag{33}$$

where in the last step we used $R(R^{-1}(\lambda)) = \lambda$. In a similar fashion we express the second term in the denominator of (29) as

$$kM_1' \left\langle M_2(x)\right\rangle_{\odot} = \left\langle M_2(x)\right\rangle_{\odot} \frac{M_1'\big(R^{-1}(\lambda)\big)}{\lambda}. \tag{34}$$

Collecting all the terms we arrive at

$$\frac{1}{\tau(k)} \sim c_1 \frac{M_1'\big(R^{-1}(\lambda)\big)}{\lambda} + c_2 \frac{M_1\big(R^{-1}(\lambda)\big)R'\big(R^{-1}(\lambda)\big)}{\lambda^2}, \tag{35}$$

where the coefficients are

$$\begin{aligned} c_1 &= 1 - \left\langle M_2(x)\right\rangle_{\odot} \\ c_2 &= -1. \end{aligned} \tag{36}$$

As we are only interested in the scaling of $\tau(k)$ with k (or λ) in the limit of large k (small λ), we can rewrite (35) without the coefficients. Indeed, taking this limit, it is the leading terms where k is raised to the highest power, which dominate the equation, providing $1/\tau(k) \sim c_1 k^a + c_2 k^b \sim k^{\max(a,b)}$, independent of c_1 and c_2. Hence, preserving only the terms relevant to the scaling, Eq. (35) becomes

$$\begin{aligned} \frac{1}{\tau(k)} &\sim \frac{1}{\lambda^2}\Big(R\left(R^{-1}(\lambda)\right) M_1'\left(R^{-1}(\lambda)\right) \\ &+ M_1\left(R^{-1}(\lambda)\right) R'\left(R^{-1}(\lambda)\right)\Big) \\ &= \frac{1}{\lambda^2}\frac{\mathrm{d}}{\mathrm{d}x}\big(M_1(x)R(x)\big)\Big|_{x=R^{-1}(\lambda)}, \end{aligned} \tag{37}$$

where, once again, we used $\lambda = R(R^{-1}(\lambda))$, allowing us to extract the prefactor of λ^{-2}. We can now write

$$\tau(k) \sim \lambda^2 Y\left(R^{-1}(\lambda)\right) \tag{38}$$

where

$$Y(x) = \left(\frac{\mathrm{d}(M_1 R)}{\mathrm{d}x}\right)^{-1}. \tag{39}$$

Equation (38) expresses $\tau(k)$ as a function of λ (12), from which its dependence on k can be obtained. It indicates that the scaling of $\tau(k)$ with k is determined directly by the dynamical functions $M_1(x)$ and $R(x)$, or, using (10), $M_1(x)$ and $M_0(x)$. Using a Hahn expansion we express (39) as

$$Y\left(R^{-1}(\lambda)\right) = \sum_{n=0}^{\infty} y_n \lambda^{\Pi_Y(n)}, \tag{40}$$

and in the limit of large k (small λ) we only keep the leading order term, namely $\lambda^{\Pi_Y(0)}$. This provides us with (38)

$$\tau(k) \sim k^{\theta}, \tag{41}$$

where

$$\theta = -2 - \Pi_Y(0). \tag{42}$$

Example - regulatory dynamics: We exemplify the use of (41) using, once again, regulatory dynamics (18). Here $M_1(x) \sim 1$, $R(x) \sim x^{-1/2}$ and $R^{-1}(x) \sim x^{-2}$, hence (39) provides

$$Y(x) = \left(\frac{\mathrm{d}}{\mathrm{d}x} x^{-\frac{1}{2}}\right)^{-1} \sim x^{\frac{3}{2}}. \tag{43}$$

The Hahn expansion (40) follows

$$Y\left(R^{-1}(\lambda)\right) \sim Y\left(\lambda^{-2}\right) \sim \lambda^{-3}, \tag{44}$$

hence $\Pi_Y(0) = -3$. Equation (42) provides us with $\theta = 1$, namely

$$\tau(k) \sim k, \tag{45}$$

in the limit of large k, as indeed we show in Supplementary Figure **1**b.

Transient observation - summary:

- The steady state activity is determined by the leading powers of (13)

$$R^{-1}(x) = \sum_{n=0}^{\infty} r_n x^{\Pi_R(n)},$$

leading to the emergence of two types of behavior

$$\begin{cases} \Pi_R(0) \neq 0 & x(k) \sim k^{\xi} & \xi = -\Pi_R(0) \\ \Pi_R(0) = 0 & x(k) \sim x_0 - k^{\xi} & \xi = -\Pi_R(1) \end{cases}$$

- The relaxation time is determined by the leading power of (40)

$$Y\big(R^{-1}(x)\big) = \sum_{n=0}^{\infty} y_n x^{\Pi_Y(n)},$$

as

$$\tau(k) \sim k^{\theta} \qquad \theta = -2 - \Pi_Y(0)$$

Supplementary Note 2. Asymptotic Observation $\mathcal{G}$

In the previous section we focused on the temporal response of the system to a local perturbation. We now turn to the asymptotic observation, $\mathcal{G}$, which describes the final perturbed state reached by the system at $t \to \infty$. To observe $\mathcal{G}$ we first allow the system (1) to reach steady state. We then induce a small permanent perturbation $\mathrm{d}x_j$ on the steady state activity of node j, forcing the system into a perturbed state in which all other node activities are shifted by $\mathrm{d}x_i$. Repeating this experiment for all nodes, one obtains the $N \times N$ response matrix

$$G_{ij} = \left| \frac{\mathrm{d}x_i/x_i}{\mathrm{d}x_j/x_j} \right| = \left| \frac{\mathrm{d}\ln x_i}{\mathrm{d}\ln x_j} \right|. \tag{46}$$

To infer the system's dynamics we focus on three functions extracted from G_{ij}, (i) stability (S_i); (ii) impact (I_i) and (iii) propagation ($\Gamma(l)$). These functions and their characteristic exponents are derived in detail in Ref. [2].Here we include a summary of the relevant results.

In Ref. [2] we show that the three functions, (i) - (iii) above, are fully determined by the power series expansion of two dynamical functions around $x = 0$. The first was already introduced in Eq. (13)

$$R^{-1}(x) = \sum_{n=0}^{\infty} r_n x^{\Pi_R(n)}, \tag{47}$$

and the second is

$$M_2\left(R^{-1}(x)\right) = \sum_{n=0}^{\infty} z_n x^{\Pi_M(n)}, \tag{48}$$

where $M_2(x)$ is taken from (1). Note, that in Ref. [2] the derivation was presented in the context of a Laurent expansion, in which the powers are integer, rather than the more general Hahn expansion, which may also include

real powers. However, the derivation of Ref. [2] readily expands (and is in fact demonstrated on) real powers, allowing us to generalize the results to Hahn series.

Stability:
We define i's stability as

$$S_i = \frac{1}{\sum_{j=1}^{N} A_{ij} G_{ij}}, \tag{49}$$

in which the denominator captures the magnitude of i's response to individual perturbations of i's nearest neighbors. If i responds strongly to neighboring perturbations, then S_i is small, indicating that node i is unstable. For all systems following (1) we show that in the limit of large k the stability of a node scales with its degree as

$$S(k) \sim k^{\delta}, \tag{50}$$

where δ is determined by the expansion of $R^{-1}(x)$ (47) as

$$\delta = \begin{cases} 0 & \Pi_R(0) \neq 0 \\ \Pi_R(1) & \Pi_R(0) = 0 \end{cases}. \tag{51}$$

Impact:
We quantify the local impact of the ith perturbation by aggregating over the response of all of i's nearest neighbors, namely

$$I_i = \sum_{j=1}^{N} A_{ij} G_{ji}. \tag{52}$$

Note that G_{ij} represents i's response to j, hence the impact (52) uses the transposed matrix G_{ji}. The impact in function of degree follows

$$I(k) \sim k^{\varphi}, \tag{53}$$

where

$$\varphi = \begin{cases} \delta - \Pi_M(0) + 1 & \Pi_M(0) \neq 0 \\ \delta - \Pi_M(1) + 1 & \Pi_M(0) = 0 \end{cases}. \tag{54}$$

Hence φ is related to the leading order terms of both (47), through δ, and (48), through Π_M.

Propagation:
To track the spread of perturbations in the network we use the distance dependent propagation function[2,8,9]

$$\Gamma(l) = \frac{1}{N}\sum_{j=1}^{N}\sum_{i\in K_j(l)} G_{ij}, \tag{55}$$

where $K_j(l)$ is the group of all nodes whose shortest path length from j equals to l. Equation (55) describes the magnitude of the perturbations experienced by *all* nodes at distance l from the source. The decay rate of $\Gamma(l)$ determines whether perturbations penetrate the network or remain localized in the source's vicinity.

In the configuration model framework, where $N \to \infty$, $P(k)$ is arbitrary, and degree correlations vanish, the network features a hyperbolic geometry, in which the number of nodes at distance l from a node grows exponentially as[4,10]

$$|K(l)| \sim e^{\alpha l}, \tag{56}$$

where

$$e^{\alpha} = \frac{\langle k^2\rangle - \langle k\rangle}{\langle k\rangle} \tag{57}$$

is the average residual degree. For networks satisfying (56), we show in Ref. [2] that

$$\Gamma(l) = e^{-\beta\alpha l} \tag{58}$$

where the *dissipation rate* β is determined by (48) through

$$\beta = \begin{cases} 0 & \Pi_M(0) \neq 0 \\ \Pi_M(1) & \Pi_M(0) = 0 \end{cases}. \tag{59}$$

up to a logarithmic correction, which depends on microscopic parameters of (1). While α is determined by the network topology, β is intrinsic to the dynamics, fully determined by the leading powers of (48).

2.1 Additional functions

The three functions described above, stability, impact and propagation, can be combined to predict other global functions pertaining to the system's response to perturbations. Indeed, all functions capturing the *global* response of the system to a perturbation are a result of the integrated effect of the local dynamics between a node and its nearest neighbors (S_i, I_i) and the patterns of propagation from there to the rest of the network ($\Gamma(l)$). Here we focus on three global functions of empirical relevance. First, we consider the probability density $P(G)$ that a randomly selected term in G_{ij} is between G and $G + \mathrm{d}G$. We show in Ref. [2] that for networks satisfying (56) we have

$$P(G) \sim G^{-\nu}, \tag{60}$$

where

$$\nu = \frac{\beta + 2}{\beta + 1}. \tag{61}$$

The second function we focus on is the *cascade* C_i associated with the perturbation of node i. It includes all nodes whose response exceeded a threshold of $G_{ji} > q$. The cascade size scales with the node's degree as

$$C(k) \sim k^{\omega}, \tag{62}$$

where

$$\omega = \frac{\beta + \varphi}{\beta + 1}. \tag{63}$$

Finally, we define the *incoming cascade* of node i as

$$Q_i = \{j | G_{ij} > q\} \tag{64}$$

namely the set of nodes j whose perturbation caused i to respond above a threshold q. As Q_i was not treated in Ref. [2] we derive it below.

Consider the response of i to a perturbation of a source node j at distance l. The tree-like structure of the configuration model network[4] predicts that there is a single path of length l connecting i to j. Hence one step before the perturbation reaches i it impacts one of i's neighbors, m, whose distance from the source j is $l - 1$. According to (58) m's response is on average

$$G_{mj} = \frac{e^{-\beta\alpha(l-1)}}{|K(l-1)|} \sim e^{-(\beta+1)\alpha(l-1)}, \tag{65}$$

where we used the fact that $\Gamma(l)$ captures the response of all $|K(l)|$ nodes at distance l from the source, and hence, on average, the response of a single node is given by $\Gamma(l)/|K(l)|$. We can now use (49) to evaluate i's response to m's perturbation as

$$G_{im} = \frac{S_i^{-1}}{k_i} \sim k_i^{-\delta-1}, \tag{66}$$

where we again used the fact that S_i^{-1} captures i's response to the perturbations of all of its k_i neighbors. Together (65) and (66) allow us to write i's response to j as

$$G_{ij} = ck_i^{-\delta-1}e^{-(\beta+1)\alpha(l-1)}, \tag{67}$$

where the constant c is independent of k_i. For j to be part of i's incoming cascade we must have $G_{ij} > q$, a condition that allows us to define the *incoming cascade radius* as

$$l_Q \sim 1 - \frac{\ln\left(qc^{-1}k^{\delta+1}\right)}{(\beta+1)\alpha} \tag{68}$$

On average, all nodes within a distance of l_Q from i will be included in Q_i. To obtain the number of nodes within this radius we follow $|K_i(l)|$, the number of nodes at distance l from i. Clearly, $K_i(1) = k_i$, after which the expansion continues as $K_i(l) \sim e^{\alpha l}$ (56). Hence the expansion from i follows

$$|K_i(l)| \sim k_i e^{\alpha(l-1)} \tag{69}$$

and the incoming cascade size becomes

$$|Q_i| \sim |K_i(l_Q)| \sim k_i \left(k_i^{-(\delta+1)}\right)^{\frac{1}{\beta+1}}, \tag{70}$$

where we omitted all terms that do not contribute to the scaling of $|Q_i|$ with k_i. Gathering all the terms we arrive at

$$|Q_i| \sim k_i^{\sigma} \tag{71}$$

where

$$\sigma = \frac{\beta-\delta}{\beta+1}. \tag{72}$$

The three additional exponents ν (61), ω (63) and σ (72) offer a redundancy, which is useful in empirical settings, where we do not always have

access to all the desired functions. For instance, measuring β involves tracking the propagation of perturbations along network paths, requiring both detailed knowledge of the network topology, as well as a high precision in measuring G_{ij}, to detect the exponentially decaying terms at large distances. However, even when these strict conditions are not satisfied, $P(G)$, which is directly extracted from G_{ij}, is still accessible. This allows us to infer β from (61), providing us access to the dissipation rate, even if we have no knowledge of the network topology at all. Similarly, measuring the cascades $C(k)$, or the incoming cascades, $Q(k)$, requires no access to A_{ij}, and the scaling exponents ω and σ can be observed even if A_{ij} is incomplete. For instance, consider the case where all we know is a random fraction p of all existing links. This would effectively result in the observed degrees being rescaled as $k_{\rm obs} \sim pk$, a rescaling which has no effect on ω and σ.

Example: regulatory dynamics: We demonstrate the predictions of $\mathcal{G}$ on regulatory dynamics (18). First we write

$$R^{-1}(x) \sim x^{-\frac{1}{a}} \tag{73}$$

and

$$M_2\left(R^{-1}(x)\right)\Big|_{x\to 0} \sim \left.\frac{x^{-\frac{h}{a}}}{1+x^{-\frac{h}{a}}}\right|_{x\to 0} \sim 1 - x^{\frac{h}{a}} + \mathcal{O}\left(x^{\frac{2h}{a}}\right). \tag{74}$$

Hence in the Hahn expansions (47) and (48) we have $\Pi_R(0) = -1/a = -2$ and $\Pi_M(0) = 0$; $\Pi_M(1) = h/a = 2/3$. As a result we predict $\delta = 0$ (51), $\varphi = 1 - h/a = 1/3$ (54) and $\beta = h/a = 2/3$ (59), from which it follows that $\nu = 8/5$ (61), $\omega = 3/5$ (63) and $\sigma = 2/5$ (72). All functions and their relevant exponents are displayed in Supplementary Figure **2**.

Supplementary Note 3. Dynamical Inference

In Supplementary Notes 1 and 2 we have shown that we can predict an array of characteristic exponents from the leading terms of $R^{-1}(x)$ (47), $Y(R^{-1}(x))$ (40) and $M_2(R^{-1}(x))$ (48). Hence we can use the observed exponents to infer the leading terms of these three functions, from which we can then reverse engineer $M_0(x)$, $M_1(x)$ and $M_2(x)$ in (1). As only the leading powers of the dynamical functions are expressed in the observed exponents, our inference will only allow us to recover (1) to its leading terms.

3.1 Inferring $R(x)$

First we use Eq. (51) to infer the structure of $R^{-1}(x)$ from δ. If $\delta = 0$ we have $R^{-1}(x)|_{x\to 0} \sim x^{\Pi_R(0)}$, with $\Pi_R(0) \neq 0$. The specific value of $\Pi_R(0)$ can be extracted from ξ using (14) to be $\Pi_R(0) = -\xi$. We can thus write $R^{-1}(x)|_{x\to 0} \sim x^{-\xi}$. Substituting $R(x)$ for the argument x we obtain

$$R^{-1}\big(R(x)\big)\Big|_{R(x)\to 0} \sim \big(R(x)\big)^{-\xi}, \tag{75}$$

which provides

$$R(x)\Big|_{x\to R^{-1}(0)} \sim x^{-\frac{1}{\xi}}. \tag{76}$$

Equation (76) provides the behavior of $R(x)$ in the limit $x \to R^{-1}(0)$, which can be either $x \to 0$ in case $\xi < 0$, or $x \to \infty$ in case $\xi > 0$. In the former case, where $x \to 0$, the solution for $R(x)$ is given by (76) with the addition of any function that involves *higher* order terms than $x^{-1/\xi}$. Indeed, such terms vanish in the limit where $x \to 0$, hence satisfying (76). Similarly, when $x \to \infty$ we have $R(x)$ following (76) up to the addition of *lower* order terms. Hence we arrive at two possible solutions, depending on the value of ξ

$$\begin{cases} R(x)\Big|_{x\to 0} = x^{-\frac{1}{\xi}} + \mathcal{O}\left(x^{\Theta_+\left(-\frac{1}{\xi}\right)}\right) & \xi < 0 \\ R(x)\Big|_{x\to \infty} = x^{-\frac{1}{\xi}} + \mathcal{O}\left(x^{\Theta_-\left(-\frac{1}{\xi}\right)}\right) & \xi > 0 \end{cases}, \tag{77}$$

where $\Theta_+(a)$ ($\Theta_-(a)$) denotes all powers greater (smaller) than a. A more compact expression can be written by introducing the operator $\Theta(a) = \Theta_{\text{sign}(a)}(a)$, in which sign(a) = a/|a|. This operator accounts for the sign of a, allowing us to write (77) as

$$R(x) \sim x^{-\frac{1}{\xi}} + \mathcal{O}\left(x^{\Theta\left(-\frac{1}{\xi}\right)}\right). \tag{78}$$

Note that while the inferred $R^{-1}(x)$ captures the behavior of the function in the limit $x \to 0$, the reconstructed $R(x)$ describes the asymptotic behavior at either $x \to 0$ or $x \to \infty$. The appropriate limit depends on the value of ξ: in case $\xi < 0$ Eq. (78) predicts $R(x)$ up to *higher* order terms, hence it becomes exact in the limit $x \to 0$; in case $\xi > 0$ the additional terms are of *lower* order and (78) becomes exact in the limit $x \to \infty$.

We now treat the case where $\delta > 0$. Equation (51) predicts $\Pi_R(0) = 0$ and $\Pi_R(1) = \delta$, hence we write $R^{-1}(x)|_{x\to 0} \sim r_0 + r_1 x^\delta$. Substituting $R(x)$ as the argument of the function we obtain

$$R^{-1}\big(R(x)\big)\Big|_{R(x)\to 0} \sim r_0 + r_1\big(R(x)\big)^\delta, \tag{79}$$

leading to

$$x\Big|_{x\to R^{-1}(0)} \sim r_0 + r_1\big(R(x)\big)^\delta. \tag{80}$$

From the structure of $R^{-1}(x)$ we have $R^{-1}(0) = r_0$, hence (80) provides

$$R(x)\Big|_{x\to r_0} \sim \left(-\frac{r_0 - x}{r_1}\right)^{\frac{1}{\delta}} + \mathcal{O}\left((r_0 - x)^{\Theta\left(\frac{1}{\delta}\right)}\right). \tag{81}$$

Note that the inference only captures the behavior of $R(x)$ in the limit $x \to r_0$, hence our degree of freedom is to add the higher powers of $(r_0 - x)$ without violating the validity of the inferred solution. Indeed, inverting (81) and taking the limit $x \to 0$ provides $\Pi_R(0) = 0$ and $\Pi_R(1) = \delta$, consistent with the observation. Finally, as the parameters r_0 and r_1 are arbitrary, we can simplify (81) by setting $r_0 = x_0$ and $-1/r_1 = 1$, providing

$$R(x) = (x_0 - x)^{\frac{1}{\delta}} + \mathcal{O}\left((x_0 - x)^{\Theta\left(\frac{1}{\delta}\right)}\right), \tag{82}$$

in which the fact that $R(x)$ is exact in the limit of $x \to x_0$ is implicit, as the higher order terms vanish in this limit. This form of dynamics in (1) describes a system with bounded activities, in which $0 < x_i < x_0$. Epidemic spreading models, for example, exhibit this form of behavior[11–13], where x_i cannot exceed unity (infected). Note that δ alone can help us distinguish between scaling dynamics ($R(x)$ follows the form (78), $\delta = 0$) and saturating dynamics ($R(x)$ follows the form (82), $\delta > 0$). The precise value of the scaling exponent in (78) however, requires us also to measure ξ.

Inferring $R(x)$ summary:
The inference of $R(x)$ requires the exponents δ and ξ, associated with the stability and steady state activities respectively:

$$R(x) \sim \begin{cases} x^{-\frac{1}{\xi}} + \mathcal{O}\left(x^{\Theta\left(-\frac{1}{\xi}\right)}\right) & \delta = 0 \\ (x_0 - x)^{\frac{1}{\delta}} + \mathcal{O}\left((x_0 - x)^{\Theta\left(\frac{1}{\delta}\right)}\right) & \delta > 0 \end{cases} . \tag{83}$$

3.2 Inferring $M_2(x)$

To obtain the structure of $M_2(x)$ we use Eq. (59). In case $\beta = 0$ we have $M_2(R^{-1}(x))|_{x\to 0} \sim x^{\Pi_M(0)}$ where $\Pi_M(0) \neq 0$. Hence we can write

$$M_2\Big(R^{-1}\big(R(x)\big)\Big)\Big|_{R(x)\to 0} \sim \big(R(x)\big)^{\Pi_M(0)}, \tag{84}$$

which leads to

$$M_2(x)\Big|_{R(x)\to 0} \sim \big(R(x)\big)^{\Pi_M(0)}, \tag{85}$$

or

$$M_2(x) \sim \big(R(x)\big)^{\Pi_M(0)} + \mathcal{O}\left(R(x)^{\Theta_+\left(\Pi_M(0)\right)}\right). \tag{86}$$

In (86) we use our degree of freedom to add higher order terms, as in the limit where $R(x) \to 0$, such terms do not harm the condition (85). Similarly to the case of $R(x)$, also here the asymptotic limit in which the inference becomes

exact is implicit in the form of the inferred $M_2(x)$. Here the appropriate limit is determined by (85) to be $R(x) \to 0$, which, according to (83) corresponds to $x \to 0$ ($\delta = 0, \xi < 0$), $x \to \infty$ ($\delta = 0, \xi > 0$) or $x \to x_0$ ($\delta > 0$). In (86) we used the fact that $\beta = 0$ to determine the functional form of $M_2(x)$, but the specific value of the exponent $\Pi_M(0)$ remains unknown. To obtain $\Pi_M(0)$ we use (54), providing

$$\Pi_M(0) = \delta + 1 - \varphi. \tag{87}$$

In case $\beta > 0$, Eq. (59) predicts $M_2(R^{-1}(x))|_{x\to 0} \sim z_0 + z_1 x^{\Pi_M(1)}$ where $\Pi_M(1) = \beta$. Writing

$$M_2\Big(R^{-1}\big(R(x)\big)\Big)\Big|_{R(x)\to 0} \sim z_0 + z_1\big(R(x)\big)^{\beta}, \tag{88}$$

we arrive at

$$M_2(x)\Big|_{R(x)\to 0} \sim z_0 + z_1\big(R(x)\big)^{\beta}, \tag{89}$$

providing

$$M_2(x) \sim z_0 + z_1\big(R(x)\big)^{\beta} + \mathcal{O}\left(R(x)^{\Theta_+(\beta)}\right). \tag{90}$$

As before the parameters z_0 and z_1 are arbitrary and can be set to any value. Indeed, our inference is aimed at predicting the powers Π_M, and not the coefficients z_n. As convention we set these parameters to be $z_0 = m_0$ and $z_1 = -1$, which describes a common scenario in which the the function $M_2(x)$ monotonically increases with $R(x)$ and saturates as $M_2(x) \to m_0$ when $R(x) \to \infty$. This form of dynamics is frequently observed in regulatory systems, *e.g.*, the Hill function in (18), in which $M_2(x)$ follows a *switch*-like behavior, bounded between 0 and $m_0 = 1$.

Inferring $M_2(x)$ summary:
The inference of $M_2(x)$ requires the exponents δ, φ and β (in addition to ξ, which is required for constructing $R(x)$):

$$M_2(x) \sim \begin{cases} \big(R(x)\big)^{\delta+1-\varphi} + \mathcal{O}\left(R(x)^{\Theta_+(\delta+1-\varphi)}\right) & \beta = 0 \\ m_0 - \big(R(x)\big)^{\beta} + \mathcal{O}\left(R(x)^{\Theta_+(\beta)}\right) & \beta > 0 \end{cases} \tag{91}$$

3.3 Inferring $M_0(x)$ and $M_1(x)$

We have now inferred the leading terms of $R(x)$ and $M_2(x)$ based on $\mathcal{G}$. This leaves us with some freedom in choosing $M_0(x)$ and $M_1(x)$, which must satisfy (10). To further narrow the inferred model space we now refer to the transient observation $\mathcal{T}$, and its additional exponent, θ, which can be connected to $M_1(x)$ through $Y(x)$ (39). First we use Eq. (42) to write

$$\Pi_Y(0) = -2 - \theta, \tag{92}$$

which in (40) provides

$$Y\left(R^{-1}(x)\right)\Big|_{x\to 0} \sim x^{-2-\theta}. \tag{93}$$

Following the same steps as in the inference of $M_2(x)$ we obtain

$$Y(x) = \big(R(x)\big)^{-2-\theta} + \mathcal{O}\left(\big(R(x)\big)^{\Theta_+(-2-\theta)}\right). \tag{94}$$

Next we use (39) to extract $M_1(x)$ as

$$\begin{aligned} M_1(x) &= \frac{1}{R(x)} \int \frac{1}{Y(x)}\, dx \\ &= \frac{1}{R(x)} \int \left[\big(R(x)\big)^{2+\theta} + \mathcal{O}\left(\big(R(x)\big)^{\Theta_+(2+\theta)}\right)\right] dx, \end{aligned} \tag{95}$$

where $R(x)$ is taken from (83). Finally, we use (10) to write

$$M_0(x) = -\frac{M_1(x)}{R(x)}, \tag{96}$$

completing the inference of all three dynamical functions.

Inferring $M_0(x)$ and $M_1(x)$ summary:
After inferring $R(x)$ (using δ and ξ), the extraction of $M_0(x)$ and $M_1(x)$ requires the exponent θ:

$$M_0(x) \sim \frac{1}{\big(R(x)\big)^2} \int \left[\big(R(x)\big)^{2+\theta} + \mathcal{O}\left(\big(R(x)\big)^{\Theta_+(2+\theta)} \right) \right] dx \quad (97)$$

$$M_1(x) \sim \frac{1}{R(x)} \int \left[\big(R(x)\big)^{2+\theta} + \mathcal{O}\left(\big(R(x)\big)^{\Theta_+(2+\theta)} \right) \right] dx. \quad (98)$$

Supplementary Note 4. Inferring Regulatory Dynamics

4.1 Numerical simulations

To test our inference formalism against regulatory dynamics we constructed a scale-free network[14] A_{ij}, with $N = 6,000$ nodes, $\langle k \rangle = 8$ and $P(k) \sim k^{-3}$. We then used a fourth order Runge-Kutta stepper to iteratively solve the N coupled equations

$$\frac{dx_i}{dt} = -Bx_i^a + \sum_{j=1}^{N} A_{ij} \frac{x_j^h}{1 + x_j^h}, \tag{99}$$

with $B = 1$, $a = 1/2$ and $h = 1/3$, until reaching steady-state for all nodes. To construct G_{ij} we induce a small permanent perturbation $\mathrm{d}x_j/x_j = 10\%$ on the activity of node j. Numerically, this is done by solving Eq. (99) for all nodes $i \neq j$. The jth equation is forced to follow

$$x_j(t) = x_j + \mathrm{d}x_j = 1.1 \times x_j. \tag{100}$$

This simulates a permanent perturbation on j, preserving the dynamical behavior of all other nodes in the network. Solving (99) coupled with (100) we allow the system to reach the new perturbed steady state, and measure $\mathrm{d}x_i$ for all nodes $i = 1, \ldots, 6,000$. This provides us with the jth column of G_{ij}. Repeating this process for all $j = 1, \ldots, 6,000$ nodes we constructed the complete G_{ij} according to (46), namely

$$G_{ij} = \frac{\mathrm{d}x_i/x_i}{\mathrm{d}x_j/x_j} = \frac{\mathrm{d}x_i/x_i}{0.1}. \tag{101}$$

For each node i we also measured the transient behavior, $x_i(t)$, from the steady state x_i to the perturbed state $x_i + \mathrm{d}x_i$, obtaining (31). We approximate the relaxation time τ_i as the time for the node to reach a fraction η

of its final perturbation, namely $x_i(t=\tau_i) = x_i + \eta \mathrm{d}x_i$. The results shown in Supplementary Figure **1**b were obtained for $\eta = 0.8$. In Supplementary Figure **3** we show results obtained for other values of η, indicating that the scaling exponent θ is not affected by the specific choice of η.

4.2 Data analysis

To obtain the characteristic exponents, both from the numerical and the empirical data (Supplementary Notes 6 and 7) we used logarithmic binning[15]. Consider the function $X(k)$, a *noisy* function which *on average* scales as $X(k) \sim k^{\alpha}$. To properly evaluate α we first we divide all nodes into W bins as

$$\mathcal{B}(w) = \left\{ i : c^{w-1} < k_i \leq c^{w} \right\} \tag{102}$$

where $w = 1, \ldots, W$ and c is constant. In (102) the wth bin includes all nodes i whose degrees k_i are between c^{w-1} and c^{w}. The parameter c is selected such that the unity of all bins $\bigcup_{w=1}^{W} \mathcal{B}(w)$ includes all nodes, hence we set $c^{W} = \max\{k_i\}$. We then plot the average degree of the nodes in each bin $k_w = \langle k_i \rangle_{i \in \mathcal{B}(w)}$ versus the average value of the function in that bin $X(k_w) = \langle X_i \rangle_{i \in \mathcal{B}(w)}$. We obtain the relevant exponent from the slope of the resulting graph, which we evaluate using common linear regression algorithms. To evaluate the measurement error for each bin we first calculated the variance in the observed function X_i across all nodes in the bin $\sigma_w^2 = \langle X_i^2 \rangle_{i \in \mathcal{B}(w)} - \langle X_i \rangle_{i \in \mathcal{B}(w)}^2$. We then set the error-bar to represent the 95% confidence interval as[16]

$$E_w = \frac{1.96 \sigma_w}{\sqrt{|\mathcal{B}(w)|}}, \tag{103}$$

where $|\mathcal{B}(w)|$ is the number of elements in $\mathcal{B}(w)$.

4.3 Inference

After measuring $\mathcal{G}$ and $\mathcal{T}$ from regulatory dynamics we find

$$\begin{array}{llll} \mathcal{G}: & \delta = 0, & \varphi = 0.33, & \beta = 0.67 \\ \mathcal{T}: & \xi = 2.0, & \theta = 1.0 & \end{array} . \tag{104}$$

From the measured values of δ and ξ we use Eq. (83) to infer

$$R(x) \sim x^{-0.5} + \mathcal{O}\left(x^{\Theta(-0.5)}\right). \tag{105}$$

Since $\beta > 0$ Eq. (91) predicts

$$M_2(x) \sim m_0 - x^{-0.33} + \mathcal{O}\left(x^{\Theta(-0.33)}\right), \tag{106}$$

where we used $(R(x))^{\beta} = x^{-0.5\times 0.67} = x^{-0.33}$. To obtain $M_1(x)$ we use (105) in (98) and take $\theta = 1$ (104), providing

$$M_1(x) \sim \left(x^{0.5} + \mathcal{O}\left(x^{\Theta_-(0.5)}\right)\right) \times \int \left(\left(x^{-0.5}\right)^3 + \mathcal{O}\left(\left(x^{-0.5}\right)^{\Theta_+(3)}\right)\right) dx, \tag{107}$$

which yields

$$M_1(x) \sim 1 + \mathcal{O}\left(x^{\Theta_-(0)}\right). \tag{108}$$

Finally, (97) provides

$$M_0(x) = \frac{M_1(x)}{R(x)} \sim \frac{1 + \mathcal{O}\left(x^{\Theta_-(0)}\right)}{x^{-0.5} + \mathcal{O}\left(x^{\Theta(-0.5)}\right)} \sim x^{0.5} + \mathcal{O}\left(x^{\Theta_-(0.5)}\right). \tag{109}$$

Hence, to leading order, the inferred model is

$$\mathbf{m} = \Big(\; x^{0.5}, \quad 1, \quad y_0 - x^{-0.33} \; \Big), \tag{110}$$

accurately capturing the leading terms of the original model (99). Note that according to the structure of the inferred $\mathbf{m}$, as given by (109), (108) and (106), the inference allows for the degree of freedom of adding *lower* order terms. Hence (110) becomes exact in the limit of $x \to \infty$, where the lower order terms become negligible (Supplementary Figure **4**).

Supplementary Note 5. Generalizing the Dynamics

The derivation of Supplementary Notes. 1 - 3 relies on the assumption that the interaction term in (1) can be factorized as $M(x_i, x_j) = M_1(x_i)M_2(x_j)$. This factorization allowed us to write, and later reverse engineer, the Hahn expansions of each function independently. In the absence of such factorization the relevant functions to infer are the projections of the mixed function $M(x_i, x_j)$, namely

$$M_1(x_i) = \frac{1}{k_i}\sum_{j=1}^{N} A_{ij}M(x_i, x_j) = \langle M(x_i, x_j)\rangle_{j\in K_i(1)} \tag{111}$$

$$M_2(x_j) = \frac{1}{k_j}\sum_{i=1}^{N} A_{ij}M(x_i, x_j) = \langle M(x_i, x_j)\rangle_{i\in K_j(1)}\,. \tag{112}$$

By averaging over the activities of the nearest neighbors of j, $x_i \in K_j(1)$, we obtain the x_i projection of $M(x_i, x_j)$, and by averaging over $x_j \in K_i(1)$ we obtain its x_j projection. Here one can no longer attribute the observables independently to the leading powers in the expansions of $M_1(x_i)$ and $M_2(x_j)$. Rather the mixed function $M(x_i, x_j)$ may lead to non-trivial dependencies between both expansions, compounding out ability to properly infer the dynamics. For instance, choosing

$$M(x_i, x_j) = x_i^{f(x_j)}, \tag{113}$$

where $f(x_j)$ is an arbitrary function of x_j, we find that the leading powers in the expansion of $M_1(x_i)$ (111) will depend on the steady state value of x_j, which may in turn depend on microscopic parameters of the system. Such dependencies violate the universality upon which our formalism builds[2], and break the derived relationships between the observable exponents and the structure of the dynamics. Still, from a practical perspective this limitation is

not very restrictive in terms of the applicability of our formalism, as functions of the form (113) are rather pathological and rarely describe the dynamics of real systems.

5.1 Diffusive dynamics - the Kuramoto model

To test our formalism against a more realistic form of non-factorizable interactions, we consider diffusive dynamics, for which the interaction term follows

$$M(x_i, x_j) = M(x_j - x_i). \tag{114}$$

As a representative example of this class of dynamics we analyze the Kuramoto model, capturing synchronization among coupled oscillators[17]. Here $x_i(t)$ represents the phase of oscillator i, which follows

$$\frac{dx_i}{dt} = w_i + \sum_{n=1}^{N} A_{in} \sin(x_n - x_i), \tag{115}$$

where w_i is i's eigen-frequency. Taking all w_i to be identical, the system reaches full synchronization, in which all oscillators lock into uniform phase. Local perturbations disturb this synchronization, allowing us to observe $\mathcal{G}$ and $\mathcal{T}$ for the coupled oscillator system. Hence we can test our formalism by attempting to reverse engineer (115), and observe the limitations and prospects of our method when treating non-factorizable dynamics.

Since we are primarily focused on the non-factorizable interaction term, we will assume, for simplicity, that the structure of $M_0(x_i) = w_i$ is known, and test how well our formalism performs against the coupling term $M(x_i, x_n) = \sin(x_n - x_i)$. The meaning of a *perturbation* in the context of oscillatory dynamics must be adapted to the nature of the system, whose components are not at a static fixed point. Hence we follow the model offered by[18]: we begin with a fully synchronized system and introduce a perturbation to node j by coupling it with an external oscillator, whose frequency is fixed at $\omega' = w_j + dw$. This is expressed in (115) by replacing the jth equation with

$$\frac{dx_j}{dt} = w_j + \sum_{n=1}^{N} A_{in} \sin(x_n - x_j) + \sin(\omega' t - x_j). \tag{116}$$

The impact of such perturbation on a specific node i is given in terms of its departure from the collective synchronization of the system. Hence denoting

the mean phase by $\langle x(t)\rangle = N^{-1}\sum_{n=1}^{N} x_n(t)$, we quantify the disturbance felt at i due to j's perturbation by the time-averaged mean square deviation

$$G_{ij} = \overline{\left(x_i - \langle x\rangle\right)^2} = \lim_{T\to\infty}\frac{1}{T}\int_0^T \left(x_i(t) - \langle x(t)\rangle\right)^2 dt. \tag{117}$$

The mean square deviation G_{ij} is a direct measure of the amplitude of the motion of i's phase with respect to the average phase, capturing the level at which j's perturbation distracts i from synchronizing with $\langle x(t)\rangle$. Perturbing all nodes as in (116) we numerically obtained $\mathcal{G}$ through (117), allowing us to extract the relevant observables. We find that for this system $\delta = \varphi = 1$ and $\nu = 2$, providing $\beta = 0$ (Supplementary Figure **5**). For these observables our formalism predicts (111) and (112) to follow

$$M_1(x_i) \sim x_0 - x_i + \mathcal{O}\left((x_0 - x_i)^{\Theta(1)}\right) \tag{118}$$

and

$$M_2(x_j) \sim y_0 - x_j + \mathcal{O}\left((y_0 - x_j)^{\Theta(1)}\right). \tag{119}$$

We are now left with two ways by which to reconstruct the coupling term of (115):

(i) Constructing a factorizable interaction

$$M(x_i, x_j) = (x_0 - x_i)(y_0 - x_j) + \dots. \tag{120}$$

(ii) By properly choosing $x_0 = \langle x_j\rangle_{j\in K_i(1)}$ and $y_0 = \langle x_i\rangle_{i\in K_j(1)}$ we can construct a non-factorizable interaction

$$M(x_i, x_j) = (x_j - x_i) + \dots. \tag{121}$$

Of the two reconstructions, option (ii) correctly captures the leading order of the actual interaction $\sin(x_j - x_i) = x_j - x_i + \mathcal{O}((x_j - x_i)^{\Theta(1)})$, indicating that our formalism is indeed capable of treating the non-factorizable (115). However, as both options (i) and (ii) are equally consistent with the observed exponents, out formalism cannot determine between them, namely we cannot use our formalism to distinguish between factorizable and non-factorizable dynamics. Such distinction must originate from prior or exogenous knowledge, *e.g.*, that the dynamics belongs to the class (114), in which case our formalism can determine that the interaction follows option (ii) above. Still, even with this limitation, the fact that our formalism is able to reduce the potential space of the dynamics to the two options (i) or (ii) provides a

significant advance on the path of reconstructing the system's microscopic dynamics. Indeed, beginning with no *a priori* knowledge about the system we used direct observation to arrive at (i) and (ii). With this at hand the problem of reverse engineering is dramatically reduced: we just need to construct a dedicated experiment/observation tailored to validating option (i) over (ii) or vice versa. Finally, we emphasize that our formalism can only be rigorously derived for factorizable dynamics. Hence, while the numerical results presented here indicate that it applies also to more general interactions, we do not know at this point how to tractably derive our reverse engineering formulation in these generalized cases.

Supplementary Note 6. Cellular Dynamics

We used high throughput microarray data obtained from Ref. [1] in which 55 transcription factors in yeast where perturbed (overexpression), and the resulting change in the expression of the remaining $N = 6,222$ target genes was measured. These perturbations, permanent in time, constitute an empirical realization of the constant perturbations described by (22). Each of the 55 perturbation experiments was conducted twice, hence the data consists of the $6,222 \times 110$ response matrix M_{ij}, in which the i, j term describes the change in the expression of gene i in response to the jth perturbation.

6.1 Pre-processing of the data

We denote the expression level of gene i in the control by $x_i(0)$, and its expression following the jth perturbation by $x_i(j)$. Using (46) the relevant response term is

$$G_{ij} = \frac{\big(x_i(j) - x_i(0)\big)/x_i(0)}{\big(x_j(j) - x_j(0)\big)/x_j(0)}. \tag{122}$$

The published data in[1], however, is given in terms of log ratios, namely

$$M_{ij} = \log_2 \left(\frac{x_i(j)}{x_i(0)} \right), \tag{123}$$

capturing the fold change in expression compared to the control. We can connect between (123) and (122) through

$$G_{ij} = \frac{2^{|M_{ij}|} - 1}{2^{|M_{\rho(j)j}|} - 1}, \tag{124}$$

where the index translator $\rho(j)$ represents the row index of the transcription factor perturbed in column j of M_{ij}. To be specific, $\rho(j)$ receives as input a

column index $j = 1, \ldots, 110$, which represents an experiment in which transcription factor X was perturbed, and provides as output the corresponding row index, $\rho(j) = 1, \ldots, 6,222$, which represents the response of the target gene X. For example in experiment $j = 1$, represented by the first column in M_{ij}, the transcription factor that was overexpressed was ABF1. The corresponding row in M_{ij} that represents the response of ABF1 to all perturbations, namely the row in which ABF1 is the target gene (YKL112W), is row number 244; hence $\rho(1) = 244$. Note that in (124), as we focus on the fold change, we use the absolute value of M_{ij}, hence G_{ij} is the same if a gene's expression was increased by a factor of 10, or decreased to 0.1 of its original expression level. Thirteen of the 55 perturbed transcription factors are not monitored in the target gene list, namely they are not part of the $6,222$ target genes, and hence $\rho(j)$ is not defined. In these experiments we do not know the size of the original perturbation and cannot properly construct (124). After eliminating these experiments from G_{ij} we are left with a $6,222 \times 84$ response matrix, following the cell's response to the perturbation of 42 transcription factors (2 experiments (columns) for each transcription factor). Approximately 12% of the entries in M_{ij} represent bad spots in the original experiment, marked in the data as $M_{ij} = \text{NaN}$. We replaced these entries with $M_{ij} = 0$, which according to (124) provides also $G_{ij} = 0$. In our data analysis we ignore these terms by focusing on $G_{ij} > 0$ only.

6.2 Observing δ and β

We measured the incoming cascades according to (64) for all targets $i = 1, \ldots, 6,222$, where we set the threshold q to $q = 0.1, 0.25, 0.5$ and 0.75. In Supplementary Figure **6**a - b we display the incoming cascade size distribution $P(Q)$ using both a log-log scale and a semi-log scale. We observe that $P(Q)$ is best fit by a bounded distribution with an exponential tail (Supplementary Figure **6**b), a finding that does not depend on the threshold q. To further test the bounded nature of $P(Q)$ we examine its exponentially decaying tail. In a bounded distribution, in the limit of large Q, $P(Q)$ features an exponential decay of the form $P(Q) \sim e^{-Q/Q_0}$, where Q_0 scales with the average incoming cascade $\langle Q \rangle$. The variance of such distributions follows $\langle Q^2 \rangle - \langle Q \rangle^2 = Q_0^2$. We can test this by setting different thresholds q in (64), where the lower is q the greater is $\langle Q \rangle$. Indeed, as shown in Supplementary Figure **6**c, the variance increases as $\langle Q^2 \rangle - \langle Q \rangle^2 \sim \langle Q \rangle^2$, providing the signature characteristic of an exponential distribution, further establishing that in this experiment $P(Q)$ is bounded. The bounded nature of $P(Q)$ indicates that Q_i does not scale with k_i, and hence $\sigma = 0$ in (71). Measuring the

distribution of terms in G_{ij} provides $P(G)$, which follows $P(G) \sim G^{-\nu}$ with $\nu = 2$ (Fig. 3a, main text). Using this result in (61) provides

$$\beta = 0. \tag{125}$$

We can now use (125) and the fact that $\sigma = 0$ in (72) to infer also that

$$\delta = 0. \tag{126}$$

6.3 Inference

We use Eq. (83) with $\delta = 0$ to predict

$$R(x) \sim x^{\eta}, \tag{127}$$

where $\eta = \Pi_R(0) \neq 0$. We then use Eq. (91) with $\beta = 0$ to obtain

$$M_2(x) \sim \big(R(x)\big)^{1-\varphi} \sim x^{\mu}, \tag{128}$$

where $\mu = \Pi_M(0) \neq 0$. Lacking access to the other exponents we cannot infer the precise value of η and μ, or the structure of $M_1(x)$. Hence the inferred model sub-space is

$$\mathbf{m} = \big(\ M_1(x)x^{\eta},\quad M_1(x),\quad x^{\mu}\ \big), \tag{129}$$

in the form of chemical reactions, as described by the law of mass action[19,20].

6.4 Validation

To test our prediction that the experiment[1] is primarily driven by protein interactions rather than regulatory interactions, we used the measured G_{ij} (124) to predict both regulatory and protein interactions, and then used the following datasets to evaluate the performance of each of these predictions.

Datasets:
The protein-protein interaction (PPI) network was downloaded from Ref. [21] . It includes $2,930$ interactions among $2,018$ genes. The relevant interactions are only ones that are represented in G_{ij}, namely interactions that involve gene pairs that are among the 42 transcription factors (columns) and $6,222$ target genes (rows) included in (124). Of the 42 transcription factors represented in G_{ij} 17 appear in the PPI dataset; similarly from the $6,222$ target genes $1,969$ are part of the PPI dataset. This allows us to construct

a sub-network of all PPIs between the 17 transcription factors and $1,969$ target genes, providing together 42 known interactions out of a potential of $17 \times 1,969 = 33,473$. The resulting sub-network has a link density of $\sim 10^{-3}$, similar to that of the original network in Ref. [21].

The regulatory interaction (RI) network[22] includes $1,079$ directed interactions among 916 genes, of which 131 are transcription factors. Focusing only on interactions represented in G_{ij}, we construct a sub-network of 40 known regulatory interactions between 37 transcription factors and 744 targets.

Data analysis:
We used G_{ij} (124) as a confidence score for the existence of a direct link between the perturbed gene in column j and the target gene i. This allowed us to construct a ranked confidence list between all pairs of nodes in the sub-networks described above: for the PPI network the list includes $17 \times 1,969 = 33,473$ entries and for the RI network it includes $37 \times 744 = 27,528$ entries. The result are two lists of the form

$$\begin{bmatrix} \text{TF} & \text{Target} & \text{Score} \\ i & j & G_{ij} \\ \vdots & \vdots & \vdots \end{bmatrix}, \tag{130}$$

in which the ranking is done according to the right column, G_{ij} (Score). If G_{ij} is a good predictor of links, the top entries should be enriched with true positives, namely i, j pairs that are among the 42 (40) PPI (RI) links. The standard measures to quantify the quality of (130) as link predictors are the area under the receiver operating curve (AUROC) and the area under the precision recall curve (AUPR)[23].

AUROC: We denote by P the overall number of true positives in (130), namely $P = 42$ for PPI and $P = 40$ for RI. The number of negatives is $N = L - P$, where L is the total length of (130), $L = 33,473$ for PPI and $L = 27,528$ for RI. We then measure the true positive rate (TPR) and the false positive rate (FPR) among the top n entries of (130) as

$$TPR(n) = \frac{TP(n)}{P} \tag{131}$$

$$FPR(n) = \frac{n - TP(n)}{N}, \tag{132}$$

where $TP(n)$ is the number of true positives among the first n entries of (130). The first quantifier, AUROC, measures the area under the curve TPR vs. FPR. For an ideal prediction, in which all true positives occupy

only the top P entries of (130) we have AUROC= 1; for a random prediction, in which the true positives are evenly distributed along the rows of (130) we have AUROC= 0.5.

AUPR: We define the precision and the recall as

$$Precision(n) = \frac{TP(n)}{n} \tag{133}$$

$$Recall(n) = TPR(n), \tag{134}$$

and measure AUPR as the area under the curve *Precision* vs. *Recall*. A perfect prediction (130) has AUPR= 1, whereas a random prediction is expected to have AUPR= P/L, which captures the density of links in the network. For PPI we have $P/L = 1.3 \times 10^{-3}$, and AUPR= 1.8×10^{-3}, representing a prediction which scores approximately 40% higher than random; for RI we have $P/L = 1.5 \times 10^{-3}$ and AUPR= 1.4×10^{-3}, indicating that G_{ij} is no better than random for predicting RIs.

Statistical significance:

To evaluate the significance of the observed AUROC and AUPR, we generated $25,000$ random lists of the form (130). Each of these lists had a random selection of P true positive edges out of a total of L potential edges, where the values of P and L match those of the relevant network, PPI ($P = 42, L = 33,473$) or RI ($P = 40, L = 27,528$). We then measured the probability density functions P(AUROC) (P(AUPR)) to obtain a certain value of AUROC (AUPR) from a random prediction. For AUROC we find that P(AUROC) follows a normal distribution (Supplementary Figure **7**a and c). This allows us to calculate the significance of the observed AUROC= x via

$$p-\text{value} = P(\text{AUROC} \geq x). \tag{135}$$

We find that for PPI, AUROC is statistically significant with p-value ~ 0.03 (exceeding the threshold $p < 0.05$). For RI, however, no statistical significance is detected as p-value ~ 0.5. For AUPR, we find that P(AUPR) has a slower decay that that of a normal distribution for large AUPR, hence using (135) overestimates the statistical significance (Supplementary Figure **7**b and d). To obtain p-value for AUPR we measured the fraction of the $25,000$ random predictions with AUPR greater or equal to the observed values. For PPI we found p-value ~ 0.07 (marginally above $p < 0.05$) and for RI p-value ~ 0.6.

Software package: All analysis and results presented in this section is included in Supplementary Software 1.

Supplementary Note 7. Human Dynamics

7.1 Data analysis

To infer the dynamics of human interactions we used two datasets: (i) **UCIonline**[24]: the data captures $\sim 6 \times 10^4$ transactions between $1,899$ users of an online instant messaging service during an $T \sim 8$ month period. (ii) **Email Epoch**[25]: this dataset monitors $\sim 3 \times 10^5$ emails sent between $3,188$ individuals over the course of $T \sim 6$ months. For both datasets we constructed the network by linking every pair of nodes where i sent at least one *package* (message/email) to j over the sampled period. While this allows for the construction of a directed network, in practice, we found that almost all links are reciprocal, to the extent that no significant difference was detected between the directed and the undirected versions of the network. The dynamics of node i is given by $x_i(t)$, denoting the number of packages sent by a user over a period of $\Delta t = 3$ hours (UCIonline) or $\Delta t = 6$ hours (Email Epoch). These choices reflect typical time scales of online messaging/email activity.

Note that in these datasets the observations $\mathcal{G}$ and $\mathcal{T}$ are not accessible directly, as we cannot induce perturbations on the activities of nodes in these systems. To overcome this limitation we used statistical measures pertaining to the nodes' activities as a proxy for the *steady-state* activity, x_i and for the response matrix G_{ij}. For the steady-state activity we used the time average of $x_i(t)$, namely

$$x_i = \frac{1}{T} \sum_{t=0}^{T} x_i(t). \tag{136}$$

As a proxy for G_{ij} we measured the correlation between the usage patterns of i and j as

$$G_{ij} = \frac{\frac{1}{T}\sum_{t=0}^{T} x_i(t)x_j(t)}{\frac{1}{T}\sum_{t=0}^{T} x_i^2(t)} = \frac{\langle x_i x_j \rangle}{\langle x_i^2 \rangle}. \quad (137)$$

If i tends to respond to j's changes in activity, this will be expressed in a large G_{ij}. The denominator is used to normalize the diagonal terms of G_{ij} to unity.

Note that the only arbitrary parameter in this analysis is Δt, which is chosen to reflect the typical time scales of email/instant messaging activity (between several hours to one day). To test if this parameter affects the results we measured $x_i(t)$, and hence (136) and (137), using three different selections for Δt. The results, shown in Supplementary Figure **8** indicate that our analysis is not sensitive to the selection of Δt.

Specific points

- In calculating $P(G)$ we threshold the G_{ij} terms, focusing only on $G_{ij} > G_{\min}$, where $G_{\min}$ represents the significance threshold. To calculate $G_{\min}$ we numerically generated 10^4 node pairs, in which each node has a random activity $x(t)$. The average activity of all nodes was set at the system's average activity, namely $\langle x(t) \rangle_t = \langle x_i \rangle_i$, taking x_i from (136), as applied to the real data. Hence the numerically generated pairs represent a randomized activity profile of the average nodes in the system. We then calculated G_{ij}^{null} for all the pairs using (137), and obtained a *null* response matrix, whose terms represent the expected response terms in the absence of activity dependencies. Response terms of the order of the terms of G_{ij}^{null} are expected to be observed from random data, and are hence insignificant. To account for this we set $G_{\min} = \langle G_{ij}^{\text{null}} \rangle + V(G_{ij}^{\text{null}})$, where $V^2(G_{ij}^{\text{null}})$ is the variance of the terms in G_{ij}^{null} (Supplementary Software 2, `PgDistribution.m`).

- The workload $W(x)$ in Email Epoch has two outliers in the limit of large k (Supplementary Figure **8**i). To calulate the scaling exponent, ζ, we ignored these two data points, which clearly violate the general trend of $W(x)$ (Supplementary Software 2, `Fig4f-jProduction.m`).

- In both Email Epoch and UCIonline the stability S_i features no detectable scaling with k_i (Supplementary Figure **8**b and g). Using the Matlab function `polyfit()` we obtain a linear slope of $\delta = 0.06$ (Email Epoch) and $\delta = -0.007$ (UCIonline). In both cases we regarded these values as insignificant, rendering δ to be zero.

- Note that our predictions for the scaling exponents are only relevant in the limit of large k. Still most measured functions feature the expected scaling for the entire range of k. The one exception is the stability, S_i, as obtained from UCIonline, where the slope tends to zero only for large k (Supplementary Figure **8**b). Hence our fit $\delta \approx 0$ is taken over the range $k \geq 10$, which we visually approximate as the point where the scaling begins (Supplementary Software 2, `Fig4a-eProduction.m`)

7.2 Independent validation: workload and node responsiveness

In the paper we derive the reverse engineered equation for human dynamics

$$\frac{dx_i}{dt} \sim -x_i^{\eta+\mu} + \sum_{j=1}^{N} A_{ij} x_i^{\mu} \left(y_0 - x_j^{-\rho}\right), \tag{138}$$

where the value of η and ρ is determined by the observed exponents (ξ, δ and ν), and μ is an arbitrary exponent. Note that while the functional form of (138) is fully inferred from the observed data, the positive and negative signs of the terms of the equation cannot be inferred from the data, and reflect a scientifically plausible selection: the first term is taken to be negative, describing a decrease in activity as packages are being sent, and hence removed from i's *to do* list. Indeed, with a negative self-dynamic term, Eq. (138) ensures that $x_i(t \to \infty) \to 0$, representing the expected outcome if no packages are received from i's neighboring nodes. The second term, taken to be positive, describes the positive contribution of i's incoming packages from its neighbors. Finally, for $M_2(x_j)$ we selected $y_0 - x_j^{-\rho}$ (and not $y_0 + x_j^{-\rho}$) to describe the increased impact of highly active neighbors (large x_j). More generally, the inferred equation has arbitrary coefficients that cannot be inferred by our formalism, namely

$$\frac{dx_i}{dt} \sim c_1 x_i^{\eta+\mu} + \sum_{j=1}^{N} A_{ij} c_2 x_i^{\mu} \left(y_0 - c_3 x_j^{-\rho}\right). \tag{139}$$

In (138) we chose $c_1 = c_3 = -1$ and $c_2 = 1$.

As explained in the main text, μ captures the level by which a node's responsiveness, R_i, depends on its activity, x_i. When μ is large, highly active nodes are significantly more responsive than non-active nodes; when it is small responsiveness is largely independent of individual activity. To evaluate μ we provide an alternative derivation for the first term of Eq. (138).

Consider the workload $W_i(t)$ experienced by node i, including all packages that have to be sent, forwarded or replied at time t. When $W_i(t)$ is large, i experiences an significant pressure to send packages, resulting in an increased activity. We express this as a scaling relationship

$$W_i(t) \sim \big(x_i(t)\big)^{\zeta}. \tag{140}$$

In this relationship we ignore the impact of i's interacting partners, which also contribute to $W_i(t)$. This is because we are only interested in deriving the first term of (138), which expresses the dynamics of $x_i(t)$ in the absence of interacting partners. Hence we consider a node that at some time t_0 accumulated a workload $W_i(t_0)$. Without any active neighbors, this node will clear its workload and relax to $x_i(t \to \infty) = 0$, as predicted by Eq. (138) if we omit the interaction term. Using (140) we obtain

$$\frac{dW_i}{dt} \sim \big(x_i(t)\big)^{\zeta-1} \frac{dx_i}{dt}. \tag{141}$$

The derivative on the l.h.s. describes the rate by which the workload changes, which is nothing else but the rate by which i sends packages. Indeed, each package sent at time t is a package extracted from $W_i(t)$, hence

$$\frac{dW_i}{dt} \sim -x_i(t). \tag{142}$$

Together with (141), Eq. (142) predicts

$$\frac{dx_i}{dt} \sim -x_i^{2-\zeta}, \tag{143}$$

which is precisely the first term of (138)

$$\frac{dx_i}{dt} \sim -x_i^{\eta+\mu}, \tag{144}$$

where

$$\zeta = 2 - \eta - \mu. \tag{145}$$

Since the value of η is provided by our inference algorithm, we can measure ζ from the empirical data, and then use (145) to infer μ, as we do in the main text.

Measuring R_i: To observe i's responsiveness, we measured all incoming mail from i's nearest neighbors j, $M(j \to i)$. We then measured all outgoing mail from i to its neighbors $M(i \to j)$. Averaging over the ratio between

incoming and outgoing mail over all of i's nearest neighbors provides the responsiveness

$$R_i = \frac{1}{k_i}\sum_{j=1}^{N} A_{ij}\frac{M(i\to j)}{M(j\to i)} = \left\langle \frac{M(i\to j)}{M(j\to i)} \right\rangle_{j\in K_i(1)}. \tag{146}$$

7.3 Practical considerations in reverse engineering

Our formalism is designed to reverse engineer a *minimal model* from the observed data. This minimal model has the form of a dynamic equation, whose structure includes only the necessary ingredients required to consistently recover the system's observed behavior. These necessary ingredients are expressed through constraints on the asymptotic behavior of the functions comprising $\mathbf{m}$, as shown in Supplementary Figure **4**. Often, however, there are additional constraints, that originate in practical considerations, that can provide further guidance in constructing the dynamic equation. To demonstrate this we consider the reverse engineered human dynamics, for which we obtained

$$M_0(x) \sim M_1(x)\Big(x^{\eta} + \mathcal{O}\left(x^{\Theta_+(\eta)}\right)\Big) \tag{147}$$

$$M_2(x) \sim y_0 - x^{-\rho} + \mathcal{O}\left(x^{\Theta_-(\rho)}\right), \tag{148}$$

where $M_1(x)$ is an arbitrary function, which we took to follow $M_1(x)\sim x^{\mu}$. The most naïve reconstruction is to take only the leading terms of (147) and (148) as we do in Eq. (18) of the main text. There are, however, several practical considerations that must be taken in order to reconstruct a realistic, *well-behaved* equation. Our first freedom is in the selection of the rate constants, namely the coefficients that precede the terms of the reverse engineered equation. This was treated above in our discussion following Eq. (139), explaining our motivation for selecting the signs of the arbitrary coefficients c_1, c_2 and c_3.

Next, we consider adding higher/lower order terms, as permitted by our formalism, to ensure the realistic behavior of the reconstructed equation. For instance, let us consider again the function $M_2(x)$ (148), which to leading order is simply

$$M_2(x) = y_0 - x^{-\rho}. \tag{149}$$

Its asymptotic behavior captures the saturation of an individual's influence on its nearest neighbors as $M_2(x \to \infty) = y_0$. The problem is that the non-asymptotic behavior of (149) is implausible from the perspective of practical human dynamics. It predicts, for small enough x that $M_2(x) < 0$, namely that if an individual's activity is lower than $x = y_0^{-1/\rho}$, his/her impact on its peers becomes negative, effectively causing them to be less active. Such behavior, may eventually lead to negative activities, and other pathological effects, that we expect the reverse engineered dynamics to avoid. Our formalism allows us to avoid these unexpected results by adding additional terms. This freedom has no bearing on the consistency of the reverse engineered model with the measured observables, as indeed our formalism makes no prediction on the behavior of $M_2(x)$ in the limit of small x, just on the asymptotic limit of large x. Still, while the observables only constrain the large x behavior of $M_2(x)$, constructing an empirically useful equation requires us to also consider the small x limit. In the case of human dynamics, we expect that $M_2(x \to 0) = 0$, describing the fact that inactive individuals have a vanishing influence on the activity of their neighbors. Hence a practical choice for $M_2(x)$ is one that satisfies both asymptotic limits, namely

$$\begin{cases} M_2(x \to 0) = 0 \\ \\ M_2(x \to \infty) = y_0 - x^{-\rho} \end{cases}, \tag{150}$$

where the later is explicitly predicted by our formalism, and the former is motivated by considering the practical realism of the reconstructed equation. A specific choice that satisfies both limits is

$$M_2(x) = \frac{x^\rho}{1 + x^\rho}, \tag{151}$$

offering to reverse engineer human dynamics as

$$\frac{dx_i}{dt} = -x_i^{\eta+\mu} + \sum_{j=1}^{N} A_{ij} x_i^\mu \frac{x_j^\rho}{1 + x_j^\rho}. \tag{152}$$

Equation (152) represents a specific dynamics out of the general class of possible models $\mathbf{m} \in \mathbb{M}(\mathcal{G}, \mathcal{T})$, which is guaranteed to predict the observed behavior, but also addresses practical issues that our formalism cannot treat by just using the observation.

To summarize, our reverse engineering formalism provides us with the asymptotic behavior in one limit, either $x \to 0$, $x \to x_0$ or $x \to \infty$ (see Supplementary Notes 3.1 and 4.3). This asymptotic behavior is sufficient for the system to recover the empirically observed behavior, and hence it

represents the only constraint provided by our formalism on the system's dynamics. To practically construct a plausible equation, however, we must often consider the anticipated behavior in the opposite limit, and use the degree of freedom allowed by our formalism to add higher/lower order terms to build a model which is sensible in both limits.

7.4 The impact of degree correlations

Our analytic derivation, presented in Supplementary Notes 1 - 3 relies on the configuration model assumption[4], that A_{ij} features negligible degree correlations. Hence although nodes may differ profoundly in terms of their degree, their neighborhood is approximately homogeneous. Some empirical networks, however, have been shown to exhibit a certain level of positive or negative degree correlations[26]. Fortunately, the observables upon which we build our formalism are rather insensitive to such microscopic properties of A_{ij}[2]. Indeed, thanks to universality, the scaling relationships we measure show little dependency on microscopic details, allowing us to relax the configuration model assumption and apply our formalism even in the presence of a certain level of degree correlations. To demonstrate this we measured the level of degree correlations in the empirical social networks that we used. Of the two networks, Email Epoch was found to exhibit a high level of degree correlations, with the average nearest neighbor degree scaling as $k_{\rm nn}(k) \sim k^{-0.6}$, namely the average degree of a node's neighborhood depends rather strongly on that node's degree (Supplementary Figure **9**a). To test if this has a noticeable effect on the observed exponents we numerically ran the reverse engineered human dynamics (152), (using $\eta = 0.79$, $\mu = 0.86$ and $\rho = 0.60$, as inferred in the main text), on the Email Epoch network, and tested whether in spite of the degree correlations, the inferred model is able to recover the observed behavior. In case the effect of degree correlations is significant it is expected that (152) will fail to reproduce the observed exponents. If, however, the model successfully recovers the observed behavior, it indicates that the observables used for reverse engineering are indeed insensitive to degree correlations, and one can reliably use our formalism in the face of empirically relevant levels of degree correlations. We find that (152) perfectly recovers the observed bahavior, showing the success of our reverse engineering formalism, and clearly indicating that degree correlations have only a negligible effect on its performance (Supplementary Figure **9**b - d).

Software package: All analysis and results presented in this section is included in Supplementary Software 2.

Supplementary References

1. G. Chua *et al.* Identifying transcription factor functions and targets by phenotypic activation. *Proc. Natl. Acad. Sci. US* **103**, 12045–50 (2006).

2. B. Barzel and A.-L. Barabási. Universality in network dynamics. *Nature Physics* **9**, 673 – 681 (2013).

3. M. Molloy and B. Reed. The size of the largest component of a random graph on a fixed degree sequence. *Combinatorics, Probability and Computing* **7**, 295–306 (1998).

4. M.E.J. Newman. *Networks - an introduction*. Oxford University Press, New York, (2010).

5. L. Schmetterer and K. Sigmund (Eds.). *Hans Hahn Gesammelte Abhandlungen Band 1/Hans Hahn Collected Works Volume 1*. Springer, Vienna, Austria, (1995).

6. U. Alon. *An Introduction to Systems Biology: Design Principles of Biological Circuits*. Chapman & Hall, London, U.K., (2006).

7. G. Karlebach and R. Shamir. Modelling and analysis of gene regulatory networks. *Nature Reviews* **9**, 770–780 (2008).

8. S. Maslov and I. Ispolatov. Propagation of large concentration changes in reversible protein-binding networks. *Proc. Natl. Acad. Sci. USA* **104**, 13655–60 (2007).

9. B. Barzel and O. Biham. Quantifying the connectivity of a network: The network correlation function method. *Phys. Rev. E* **80**, 046104–15 (2009).

10. D. Krioukov, F. Papadopoulos, M. Kitsak, A. Vahdat and M. Boguñá. Hyperbolic geometry of complex networks. *Physical Review E* **82**(3), 036106 (2010).

11. L. Hufnagel, D. Brockmann and T. Geisel. Forecast and control of epidemics in a globalized world. *Proc. Natl. Acad. Sci. US* **101**, 15124–9 (2004).

12. P.S. Dodds and D.J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology* **232**, 587–604 (2005).

13. R. Pastor-Satorras and A. Vespignani. Epidemic spreading in scale-free networks. *Phys. Rev. Lett.* **86**, 3200–3203 (2001).

14. A.L. Barabási and R. Albert. Emergence of scaling in random networks. *Science* **286**, 509 – 512 (1999).

15. S. Milojević. Power law distributions in information science: Making the case for logarithmic binning. *Journal of the American Society for Information Science and Technology* **61**(12), 2417–2425 (2010).

16. D.R. Cox and D.V. Hinkley. *Theoretical Statistics*. Chapman & Hall, (1974).

17. Y. Kuramoto. *Chemical oscillations, waves and turbulance*. Springer-Verlag Berlin, Heidelberg, (1984).

18. D.H. Zanette. Disturbing synchronization: Propagation of perturbations in networks of coupled oscillators. *Europhys. Lett.* **68**, 356 (2004).

19. J.D. Murray. *Mathematical Biology*. Springer, Berlin, (1989).

20. E.O. Voit. *Computational Analysis of Biochemical Systems*. Cambridge University Press, New York, NY, (2000).

21. H. Yu *et al.* High-quality binary protein interaction map of the yeast interactome network. *Science* **322**, 104–110 (2008).

22. R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii and U. Alon. Network motifs: Simple building blocks of complex networks. *Science* **298**, 824–827 (2002).

23. D. Marbach, J.C. Costello, R. Küffner, N.M. Vega, R.J. Prill, D.M. Camacho, K.R. Allison, The DREAM5 Consortium, M. Kellis, J.J. Collins and G. Stolovitzky. Wisdom of crowds for robust gene network inference. *Nature Methods* **9**, 796–804 (2012).

24. T. Opsahl and P. Panzarasa. Clustering in weighted networks. *Social Networks* **31**, 155–163 (2009).

25. J.-P. Eckmann, E. Moses and D. Sergi. Entropy of dialogues creates coherent structures in e-mail traffic. *Proc. Natl. Acad. Sci. US* **101**, 14333–7 (2004).

26. M.E.J. Newman. Assortative mixing in networks. *Phys. Rev. Lett.* **89**, 208701 (2002).