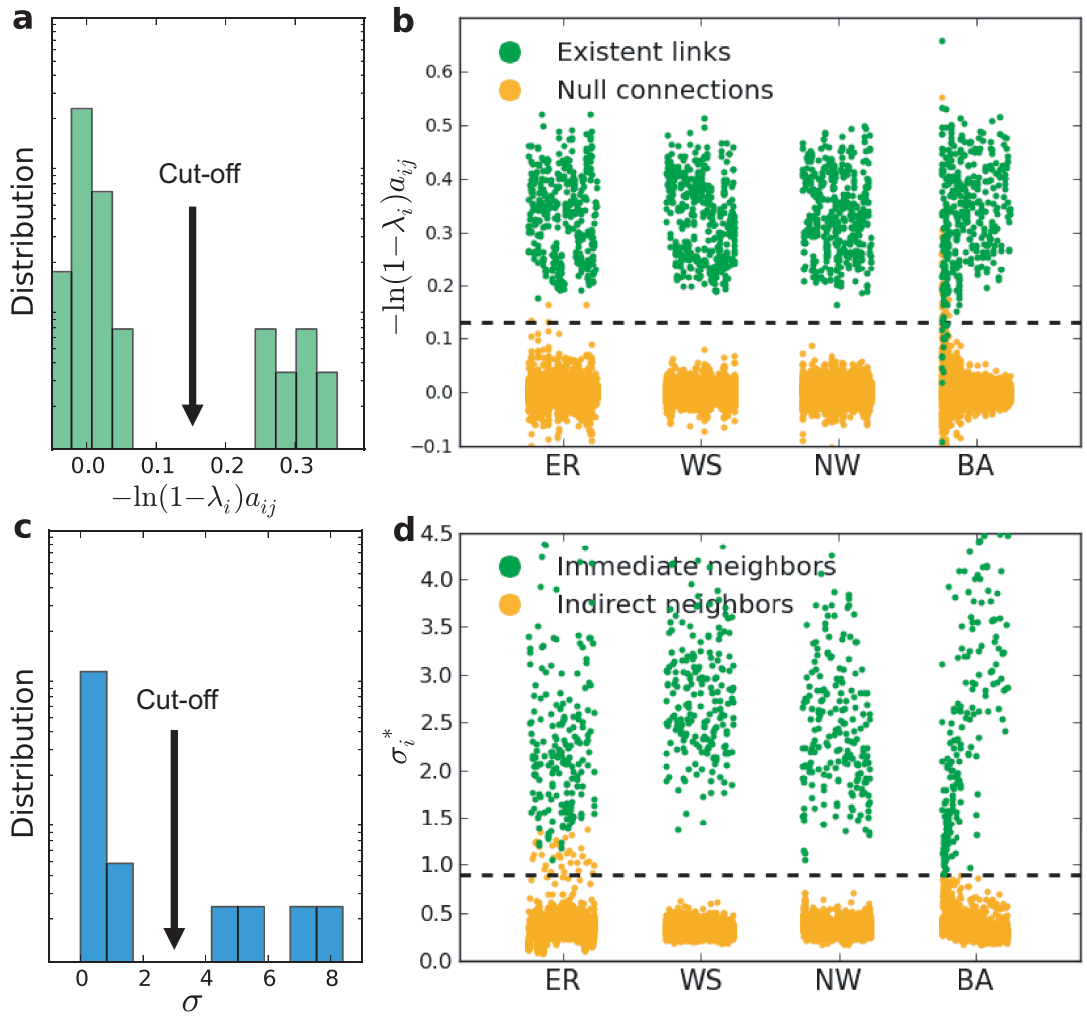
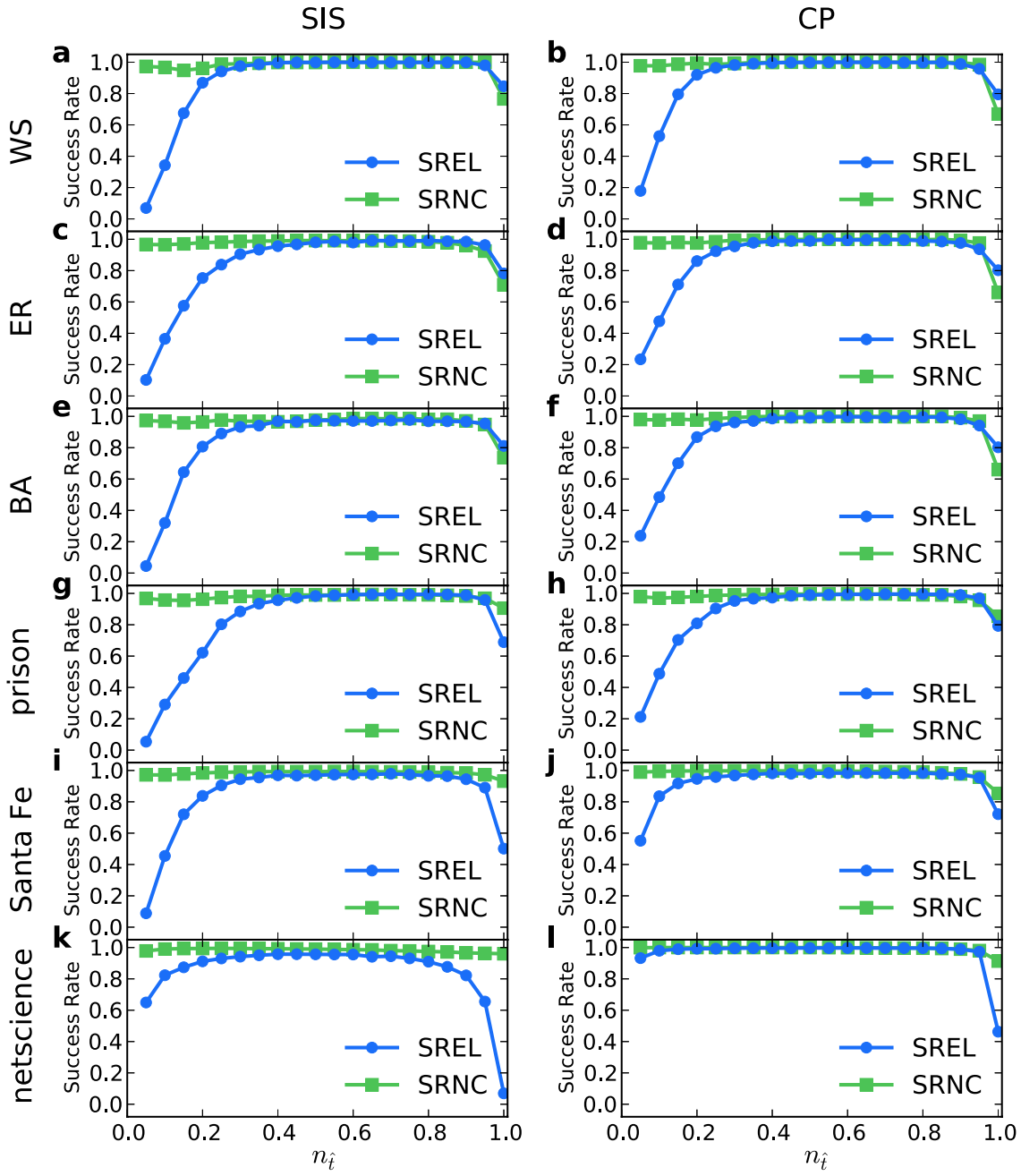
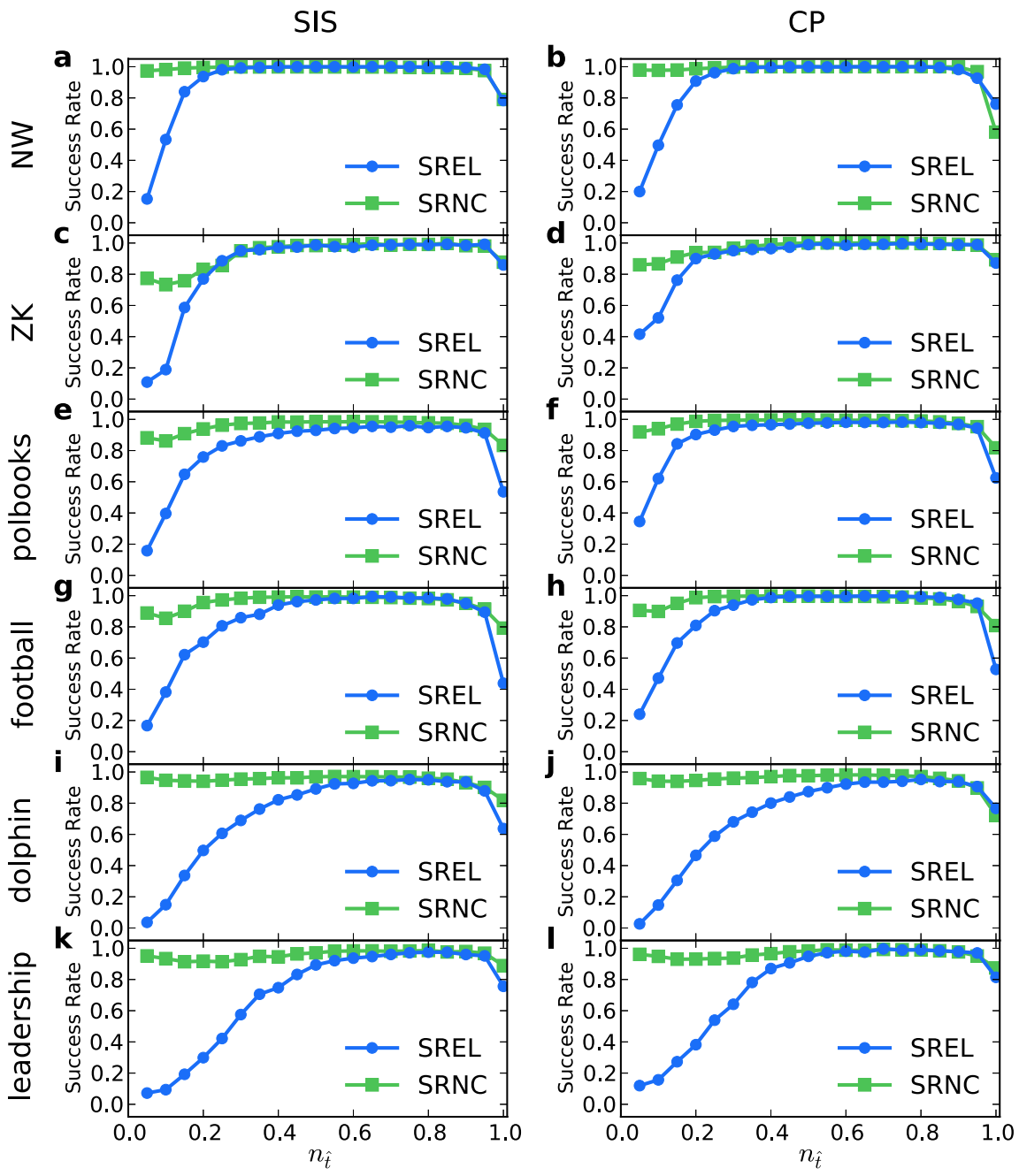
Supplementary Figures



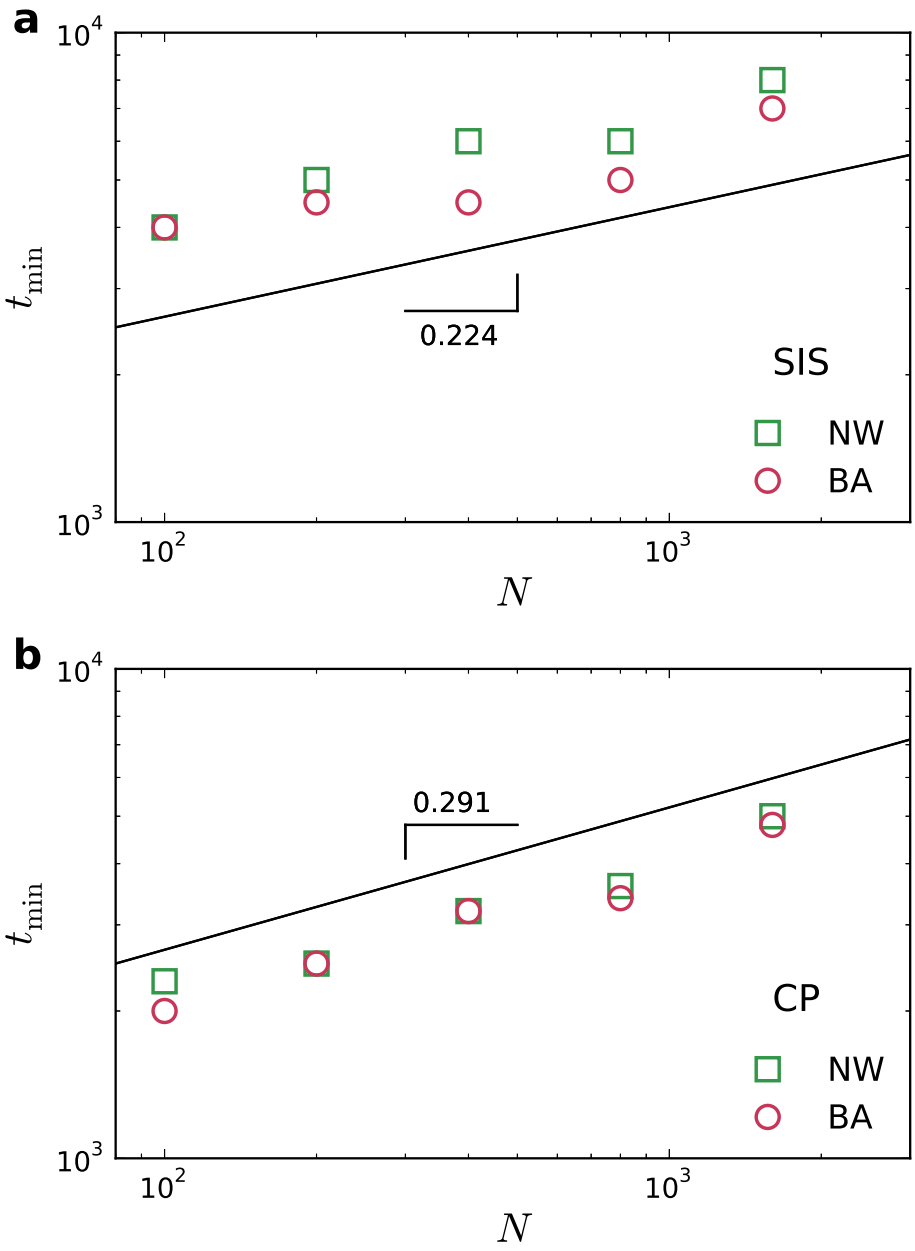
Supplementary Figure 1 | **Setting cut-off to identify links and hidden source.** (**a**) Coarse-grained distribution of element values $\ln(1-\lambda_i)a_{ij}$ times -1 in the reconstructed vector $\mathbf{X}$ of all nodes in a Newman-Watts small-world network (NW), where bin accumulation is implemented for the original distribution. The peak centered at zero corresponds to null connections and the other corresponds to actual links. A cut-off value can be set within the gap between the two peaks. (**b**) The values of $\ln(1-\lambda_i)a_{ij}$ times -1 in the reconstructed vector $\mathbf{X}$ for Erdös-Rényi random networks (ER), Watts-Strogatz small-world networks (WS), NW and Barabási-Albert scale-free networks (BA), where green and yellow points correspond to existent links and null connections, respectively. All the networks share the same cut-off value of about 0.13. (**c**) Coarse-grained distribution of the structural variance σ in the NW network, where bin accumulation is exercised on the original distribution. The largest peak centered close zero corresponds to nodes of indirect neighbors of the hidden source and the large positive values of σ are associated with the immediate neighbors of the hidden source. A cut-off can be set within the largest gap. (**d**) The variance σ_i^* of element values in the reconstructed vector $\mathbf{x}$ of all nodes in the presence of a hidden source for ER, WS, NW and BA networks, where green and yellow points correspond to immediate neighbors and indirect neighbors of the hidden source, respectively. All the networks share the same cut-off values of about $\sigma_i^* = 0.9$. The network size in **a,b** is 200 and in **c,d** is 50, $\Theta = 0.25$, $\Delta = 0.45$, λ_i and δ_i are uniformly distributed in the ranges $(0.2, 0.4)$ and $(0.4, 0.6)$, respectively. SIS dynamics are used for the illustration. Panels **a** and **c** correspond to Fig. 2a with $n_{\hat{i}} = 0.4$ and Fig. 5 in the main text, respectively.



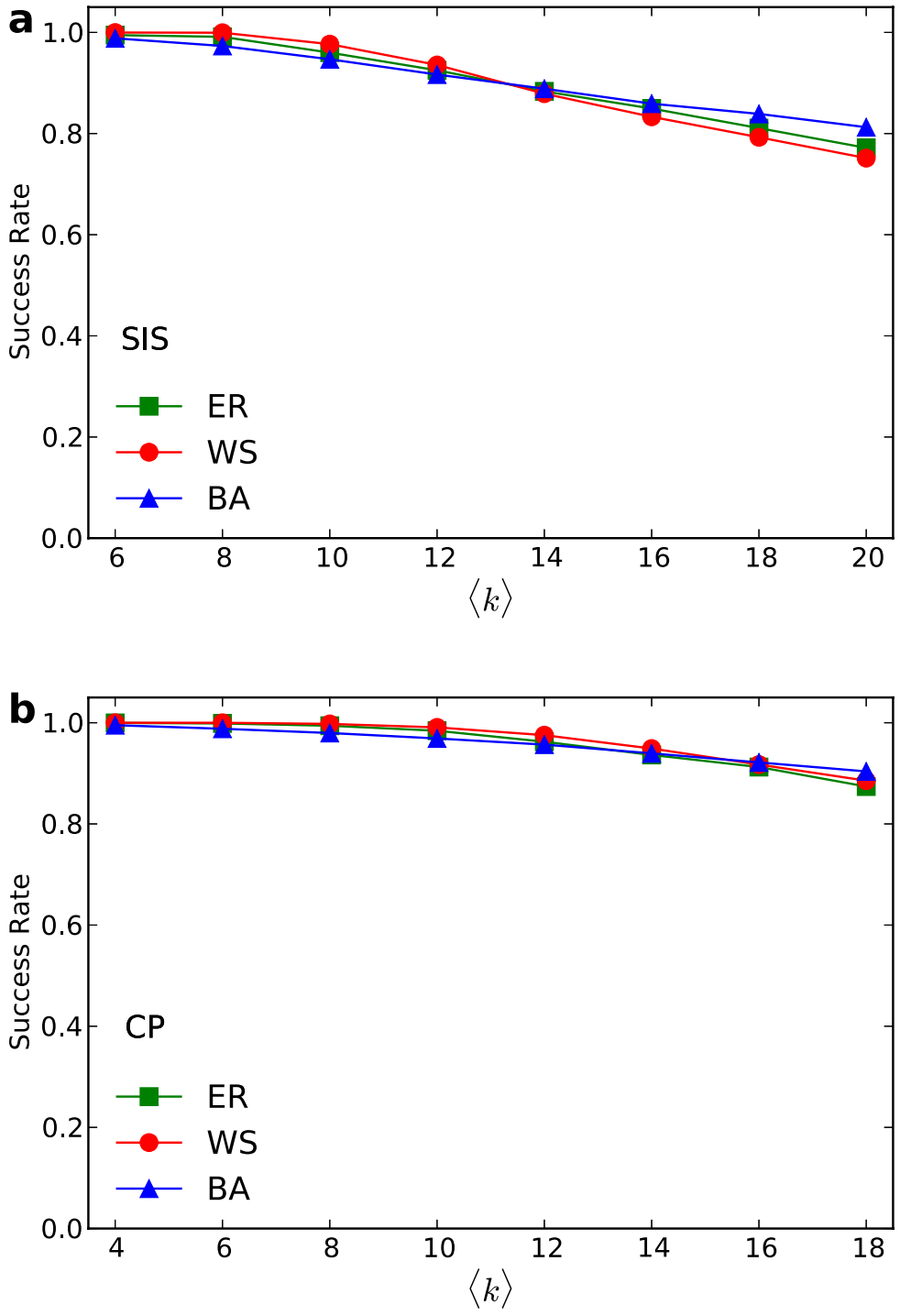
Supplementary Figure 2 | Reconstruction accuracy as a function of the fraction $n_{\hat{t}}$ of base strings. Success rates of SREL and SRNC as a function of $n_{\hat{t}}$ for SIS and CP dynamics in combination with (**a**-**b**) WS networks, (**c**-**d**) ER networks, (**e**-**f**) BA networks, (**g**-**h**) a prison inmate network (prison), (**i**-**j**) the Santa Fe institute relationship network (Santa Fe) and (**k**-**l**) the collaboration network of network science (netscience). For SIS dynamics, $\Theta = 0.25$, $\Delta = 0.45$, λ_i and δ_i are distributed in the ranges $(0.2, 0.4)$ and $(0.4, 0.6)$, respectively. For CP dynamics, $\Theta = 0.35$, $\Delta = 0.45$, λ_i and δ_i are distributed, respectively, in the ranges $(0.7, 0.9)$ $(0.2, 0.4)$. The results are obtained from ensemble average over 10 independent realizations. The length of time series is 5×10^4 for all the networks. For model networks, $N = 100$ and $\langle k \rangle = 4$. The network sizes of real networks are listed in Supplementary Table 1.



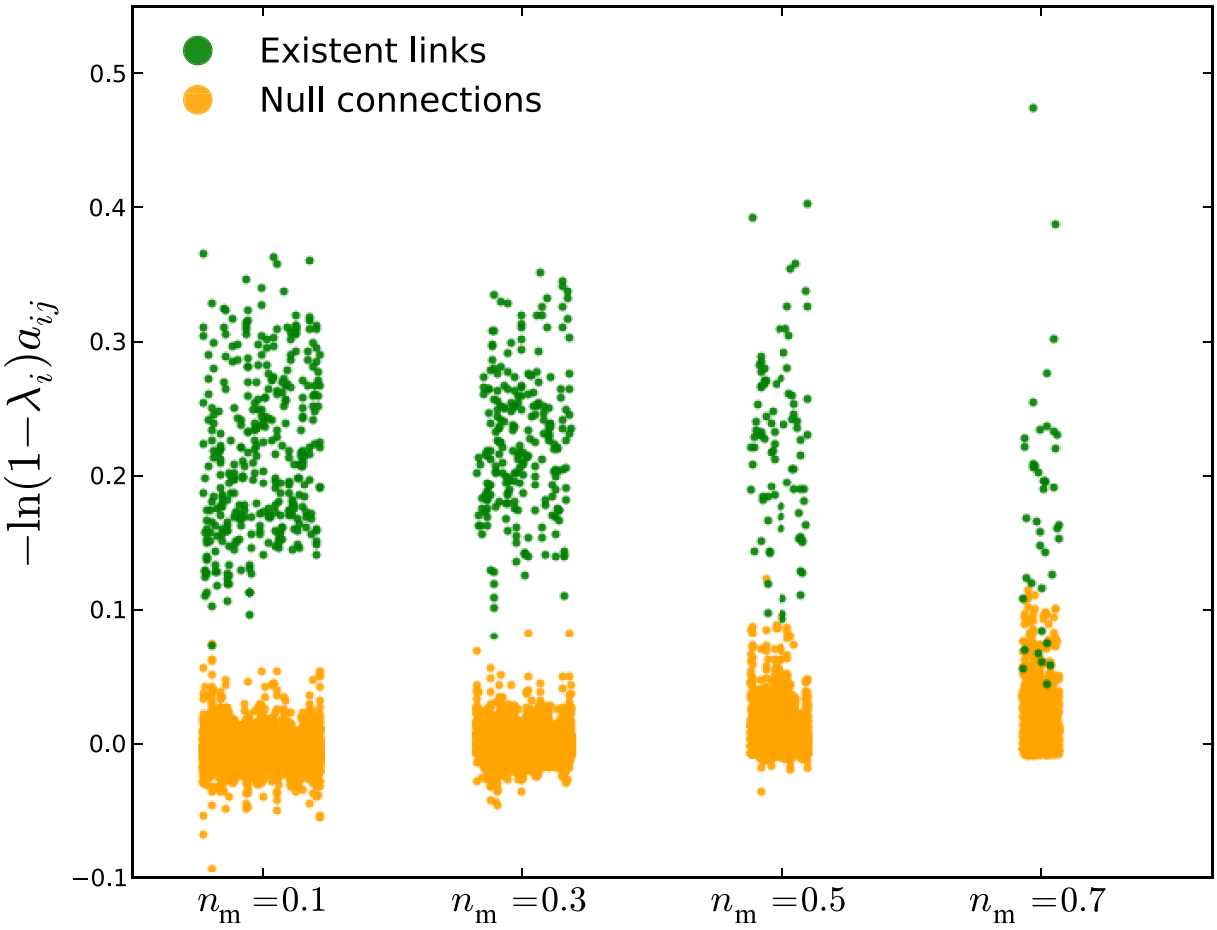
Supplementary Figure 3 | Reconstruction accuracy as a function of the fraction $n_{\hat{t}}$ of base strings. Success rates of SREL and SRNC as a function of $n_{\hat{t}}$ for SIS and CP dynamics in combination with (**a**-**b**) NW networks, (**c**-**d**) Zachary's karate club network (ZK), (**e**-**f**) political book network (polbooks), (**g**-**h**) American College football network (football), (**i**-**j**) dolphin social network network (dolphin) and (**k**-**l**) leadership network (leadership). The parameters are the same as in Supplementary Fig. 2. For NW network, $N = 100$ and $\langle k \rangle = 4$. The network sizes of real networks are listed in Supplementary Table 1.



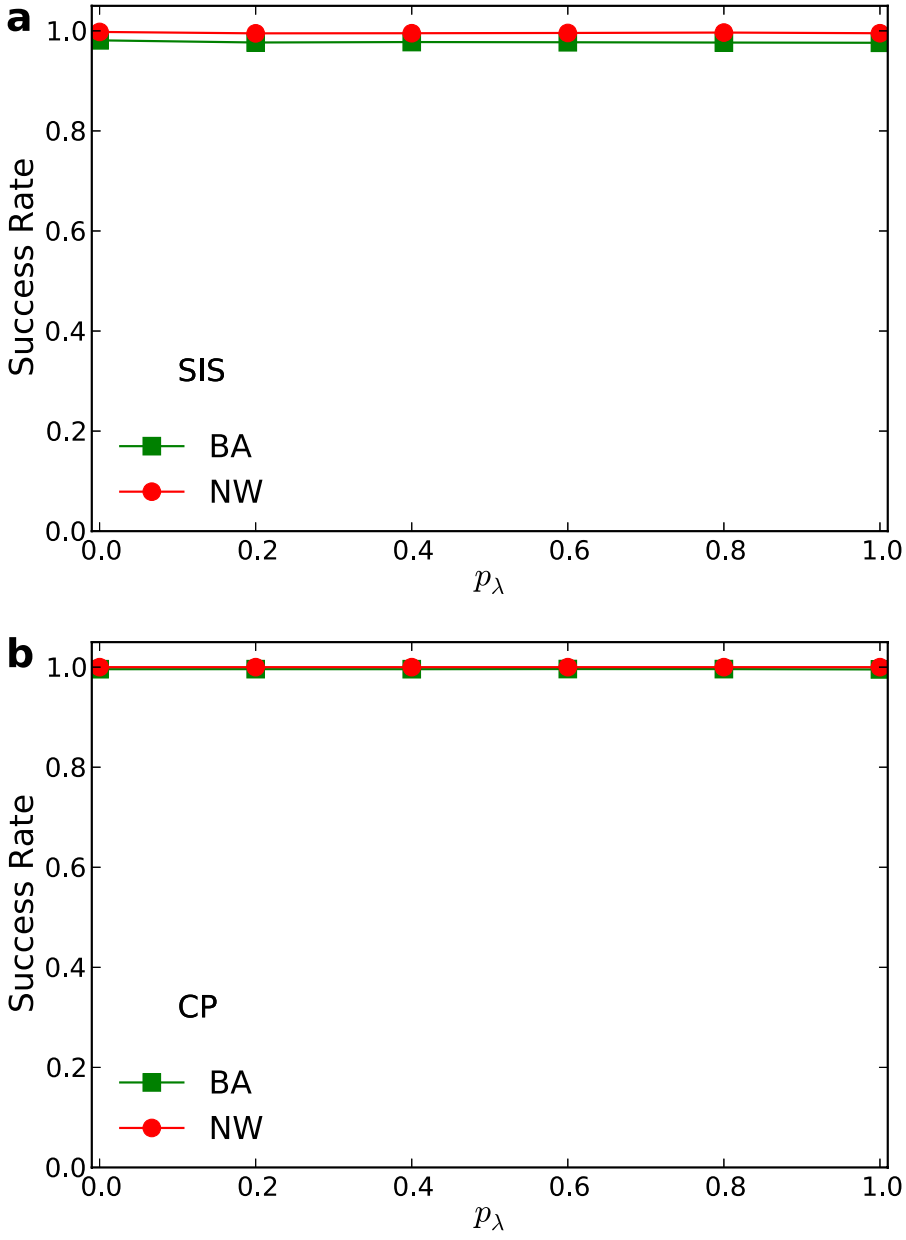
Supplementary Figure 4 | Minimum length of time series versus network size. The minimum absolute length $t_{\min}$ required for 95% success rates of reconstruction as a function of network size N for (**a**) SIS and (**b**) CP dynamics taking place on NW and BA networks, respectively. Lines are for visual guide. The average node degree is $\langle k \rangle = 4$. Other parameters are the same as in Supplementary Fig. 2. The relative minimum length $n_t^{\min}$ is shown in Fig. 3c,d in the main text.



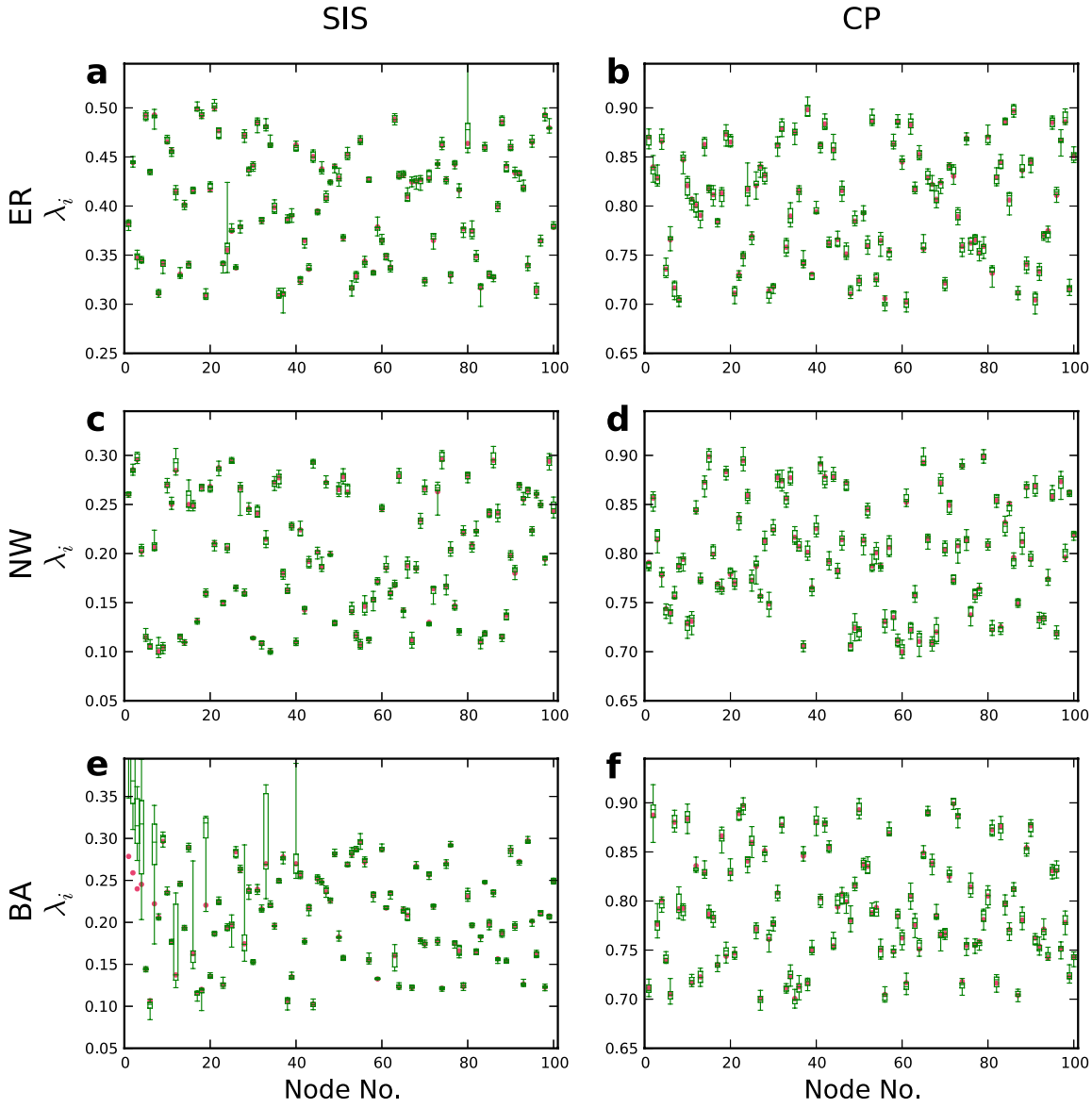
Supplementary Figure 5 | Effect of average degree on reconstruction accuracy. Success rates as a function of average degree $\langle k \rangle$ for (**a**) SIS and (**b**) CP dynamics on ER, NW and BA networks. The network size N is 200 and other parameters are the same as in Supplementary Fig. 2.



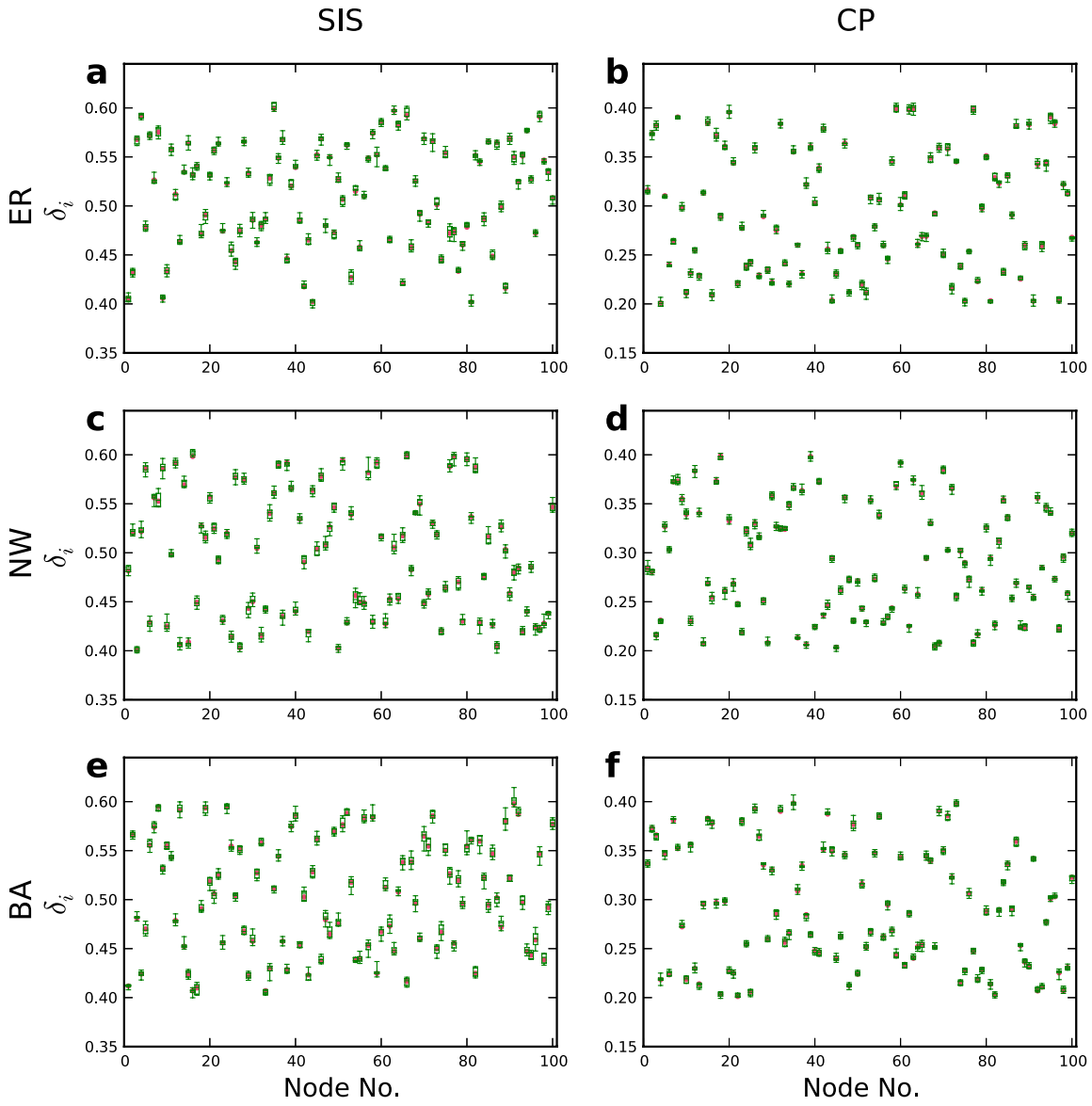
Supplementary Figure 6 | Effect of missing time series on reconstructed vector X. Element value $\ln(1-\lambda_i)a_{ij}$ times -1 for different fraction n_{m} of nodes that are externally inaccessible. The nodes for which time series are missing and their links are removed from the network. NW networks of size $N = 100$ are used to illustrate the effect of missing time series. Other parameters are the same as in Supplementary Fig. 2.



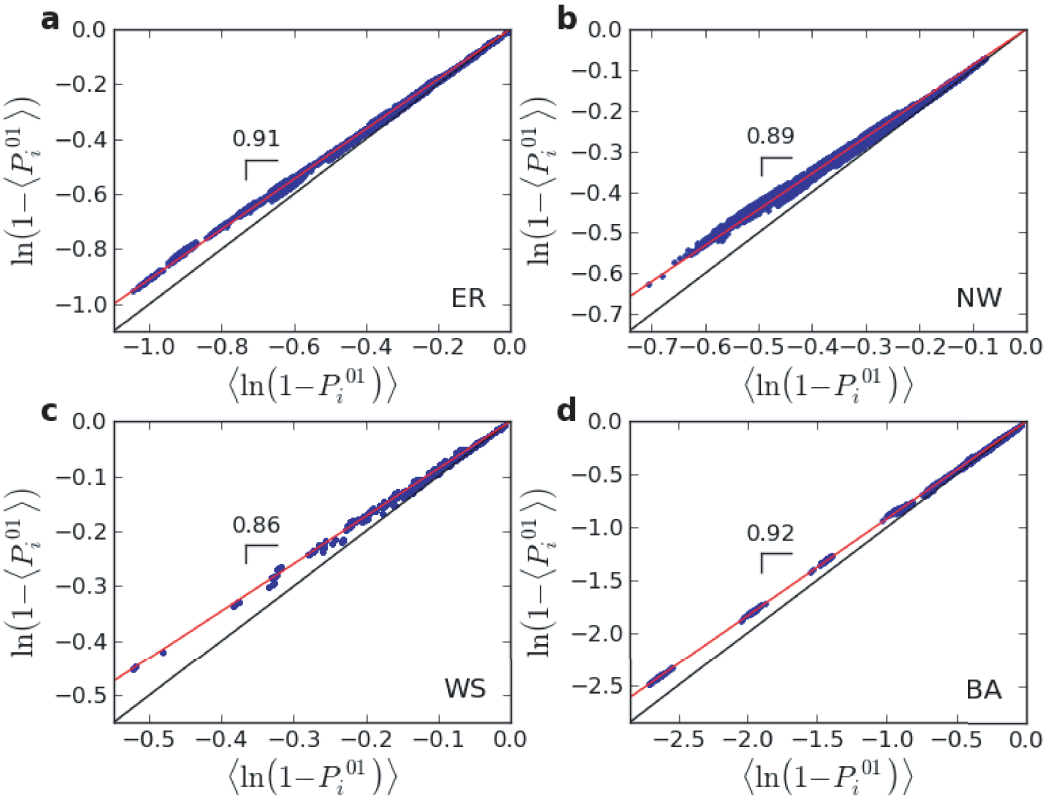
Supplementary Figure 7 | Reconstruction accuracy with respect to time-variant infection rates. Success rate as a function of the probability p_λ of changing the infection rates of individuals at each time step for SIS and CP taking place on BA and NW networks. With probability p_λ, the infection rate of a node is randomly changed within the range $(0.1, 0.3)$. The network size N is 500 and other parameters are the same as in Supplementary Fig. 2.



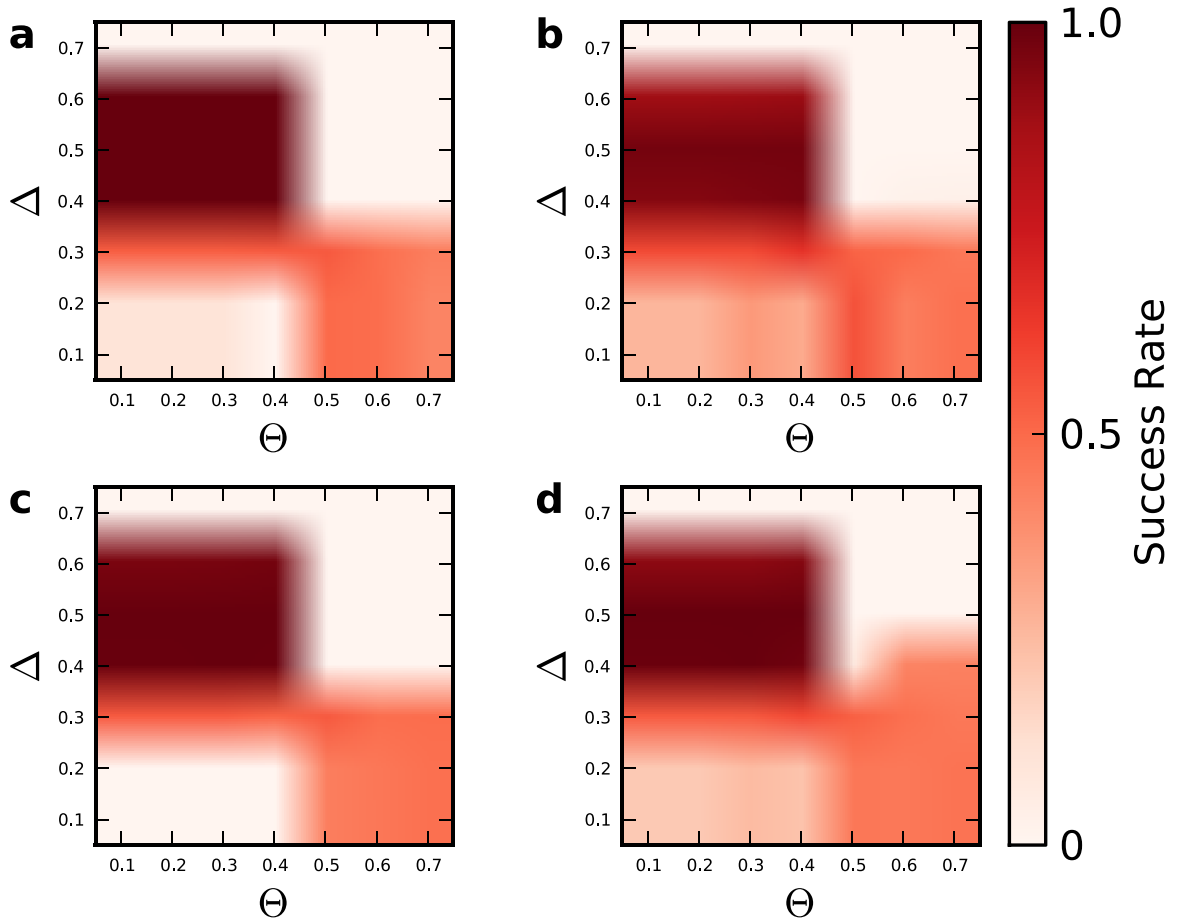
Supplementary Figure 8 | Inference of inhomogeneous infection rates. Box-plots of differences between actual and predicted infection rates λ_i for SIS and CP dynamics in combination with (**a**-**b**) ER, (**c**-**d**) NW and (**e**-**f**) BA networks. The network size N is 100 and the threshold Θ and Δ are the same as in Supplementary Fig. 2. Red circles denote the actual values, the upper and lower bars outside the boxes represent 9% and 91% of the rank of predicted values, respectively, the upper and lower bound of the boxes represent 25% and 75% of the rank of predicted values, respectively, and the bars inside the boxes characterize the median value. The results are obtained from 10 independent realizations.



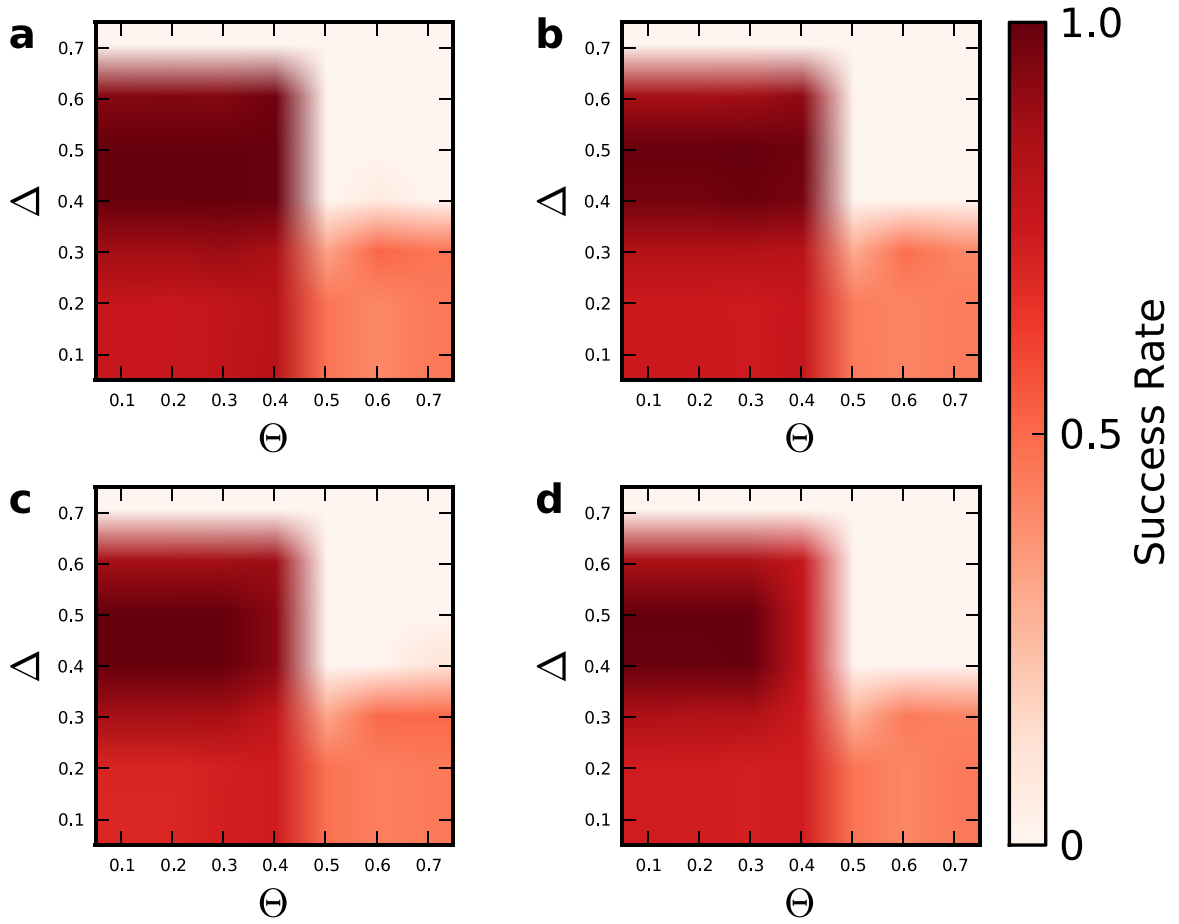
Supplementary Figure 9 | Inference of inhomogeneous recovery rates. Box-plots of differences between actual and predicted recovery rates δ_i for SIS and CP dynamics in combination with (**a**-**b**) ER, (**c**-**d**) NW and (**e**-**f**) BA networks. The network size N is 100 and the values of thresholds Θ and Δ are the same as in Supplementary Fig. 2. The standard box-plots have the same meaning as that in Supplementary Fig. 8.



Supplementary Figure 10 | Difference between $\ln(1-\langle P_i^{01}\rangle)$ and $\langle\ln(1-P_i^{01})\rangle$ for SIS dynamics. The two quantities for (**a**) ER, (**b**) NW, (**c**) WS and (**d**) BA networks. The network size N is 100 and other parameters are the same as in Supplementary Fig. 2. The red line is the best fit of data and the black line is the diagonal line. The slope of each fitting is marked.



Supplementary Figure 11 | Dependence of success rates on thresholds for SIS dynamics. success rates in the Δ-Θ parameter plane for SIS dynamics on (**a**) WS, (**b**) BA, (**c**) NW and (**d**) ER networks. The network size N is 100 and $\langle k \rangle = 4$. The infection and recovery rates of nodes are the same as in Supplementary Fig. 2. The results are obtained from 10 independent realizations.



Supplementary Figure 12 | Dependence of success rates on thresholds for CP dynamics. Success rates in the Δ-Θ parameter plane for CP dynamics on (**a**) WS, (**b**) BA, (**c**) NW and (**d**) ER networks. The network size N is 100 and $\langle k \rangle = 4$. The infection and recovery rates of nodes are the same as in Supplementary Fig. 2. The results are obtained from 10 independent realizations.

Supplementary Table

Supplementary Table 1 | Description of the real social networks analyzed in the paper. N and L denote the total numbers of nodes and links, respectively.

Name	N	L	Description
ZK[1]	34	78	Social network of friendships of a karate club
Santa Fe [2]	118	200	Scientific collaboration network of the Santa Fe Institute
leadership [3]	32	96	College students in a course about leadership
prison [3, 4]	67	182	Social networks of positive sentiment (prison inmates)
dolphin [5]	62	159	social network of dolphins
polbooks [6]	105	441	A network of books about US politics
football [2]	115	613	The network of American football games, Fall 2000.
netscience [7]	1589	2742	A coauthorship network of scientists working on networks.
netscience* [7]	379	914	The largest component of netscience.

Supplementary Note 1: Compressed sensing theory and algorithm

The basic idea of compressed sensing is to reconstruct sparse data or a signal from a few observations whose number is much less than that of the original data sets [8, 9, 10, 11]. The observations are measured by linear projections of the original data on a few predetermined, random vectors. Since the requirement for observations is considerably less comparing to that in conventional signal reconstruction schemes, compressed sensing has been developed into a powerful technique to obtain high-fidelity signal, especially in cases where sufficient observations are not available.

Restricted Isometry Property. To reconstruct vector $\mathbf{X}$ in the form $\mathbf{Y} = \mathbf{\Phi} \cdot \mathbf{X}$ by using the compressed sensing algorithm, the matrix Φ needs to satisfy Restricted Isometry Property (RIP) [9, 10, 11]. Specifically, matrix $\mathbf{\Phi}$ is said to satisfy RIP if

$$(1 - \varepsilon_s)\|\mathbf{X}\|_2^2 \leq \|\mathbf{\Phi X}\|_2^2 \leq (1 + \varepsilon_s)\|\mathbf{X}\|_2^2, \tag{1}$$

holds simultaneously for all s-sparse vectors $\mathbf{X}$, for some $\varepsilon_s \in (0, 1)$. Prior works have verified that RIP holds with high probability for sampling matrices $\mathbf{\Phi}$ whose entries are independently and identically distributed (i.i.d.) realizations of certain random variables, provided that the number of rows in the matrix is large enough. For our reconstruction problem, the randomness of matrix $\mathbf{\Phi}$ is induced by the stochastic effect in the spreading process, allowing us to successfully reconstruct the neighboring vector $\mathbf{X}$ from a small number of base strings using compressed sensing algorithm. The accurate reconstruction results obtained from a variety of complex networks suggest that the condition of RIP is met, due to the intrinsic stochastic nature of the spreading dynamics. With respect to data requirement, our experience suggests that the main factor determining the amount of data required is the sparsity of vector $\mathbf{X}$, where a sparser vector $\mathbf{X}$ requires less data for reconstruction. Nevertheless, provided with sufficient data, one can recover even relatively dense vectors.

Codes. The Matlab codes of the compressed sensing algorithm [12, 13] based on the L1-norm optimization that we used in this paper can be downloaded from [14].

Supplementary Note 2: Reconstruction framework of CP dynamics

For the CP dynamics, the treatment is quite similar to that of SIS dynamics. Specifically, averaging over time t restricted by equation (14) on both sides of equation (7), we obtain

$$\langle P_i^{01}(t)\rangle = \frac{\lambda_i}{k_i}\sum_{j=1,j\neq i}^{N} a_{ij}\langle S_j(t)\rangle. \tag{2}$$

Substituting $\langle P_i^{01}(\hat{t}_\alpha)\rangle$ by $\langle S_i(\hat{t}_\alpha+1)\rangle$, we obtain equation (8). According to equation (16), a set of base strings at different time $\hat{t}$ can be identified. For each base string, repeating the above process gives rise to a set of equations at different time $\hat{t}_\alpha$ ($\alpha = 1, \cdots, m$) that can be expressed in the matrix form $\mathbf{Y}_{m\times 1} = \mathbf{\Phi}_{m\times(N-1)} \cdot \mathbf{X}_{(N-1)\times 1}$:

$$\begin{bmatrix} \langle S_i(\hat{t}_1+1)\rangle \\ \langle S_i(\hat{t}_2+1)\rangle \\ \vdots \\ \langle S_i(\hat{t}_m+1)\rangle \end{bmatrix} = \begin{bmatrix} \langle S_1(\hat{t}_1)\rangle & \cdots & \langle S_{i-1}(\hat{t}_1)\rangle & \langle S_{i+1}(\hat{t}_1)\rangle & \cdots & \langle S_N(\hat{t}_1)\rangle \\ \langle S_1(\hat{t}_2)\rangle & \cdots & \langle S_{i-1}(\hat{t}_2)\rangle & \langle S_{i+1}(\hat{t}_2)\rangle & \cdots & \langle S_N(\hat{t}_2)\rangle \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \langle S_1(\hat{t}_m)\rangle & \cdots & \langle S_{i-1}(\hat{t}_m)\rangle & \langle S_{i+1}(\hat{t}_m)\rangle & \cdots & \langle S_N(\hat{t}_m)\rangle \end{bmatrix} \begin{bmatrix} a_{i1}/k_i \\ \vdots \\ a_{i,i-1}/k_i \\ a_{i,i+1}/k_i \\ \vdots \\ a_{iN}/k_i \end{bmatrix}. \tag{3}$$

Compared with the reconstruction of SIS dynamics, the matrix $\mathbf{\Phi}$ has the same form, but the vectors $\mathbf{Y}$ and $\mathbf{X}$ are different. In fact, reconstruction of CP is expected to be more accurate because fewer approximations are used in the derivation of the reconstruction form. For example, the derivation of equation (15) from equation (14) with approximation in the main text for SIS dynamics is absent for CP.

Supplementary Note 3: Finding base strings for reconstruction

We provide details for finding base strings and other strings derived from these strings. Given a set of time series of length t, to reconstruct the neighboring vector $\mathbf{X}$ of an arbitrary node i, we first find all possible base strings according to the threshold Θ. The procedure is as follows. Randomly choose a time t and set the string $S_{-i}(t)$ to be the first base string $S_{-i}(\hat{t}_1)$. Then find the second base string $S_{-i}(\hat{t}_2)$ that satisfies $H[S_{-i}(\hat{t}_1), S_{-i}(\hat{t}_2)] > \Theta$. The third base string $S_{-i}(\hat{t}_3)$ should satisfy $H[S_{-i}(\hat{t}_1), S_{-i}(\hat{t}_3)] > \Theta$ and $H[S_{-i}(\hat{t}_2), S_{-i}(\hat{t}_3)] > \Theta$. Repeat this process until no more base strings can be found, which gives rise to a set of base strings $S_{-i}(\hat{t}_1), S_{-i}(\hat{t}_2), \cdots, S_{-i}(\hat{t}_m)$, for which the normalized Hamming distance between each pair of base strings is large than Θ.

For a base string $S_{-i}(\hat{t}_\alpha)$, we enumerate the time series to pick all strings $S_{-i}(t_\nu)$ that satisfies $H[S_{-i}(\hat{t}_\alpha), S_{-i}(t_\nu)] < \Delta$, which gives a set of t_ν ($\nu = 1, 2, \cdots, l$). We can then calculate two average values based on $\hat{t}_\alpha$ and t_ν:

$$\begin{aligned} \langle S_i(\hat{t}_\alpha + 1)\rangle &= \frac{S_i(\hat{t}_\alpha + 1) + \sum_{\nu=1}^{l} S_i(t_\nu + 1)}{l+1}, \\ \langle S_j(t_\alpha)\rangle &= \frac{S_j(\hat{t}_\alpha) + \sum_{\nu=1}^{l} S_j(t_\nu)}{l+1}, \quad (j = 1, 2, \cdots, i-1, i+1, \cdots, N). \end{aligned} \tag{4}$$

The two averaged quantities are used to build the vector $\mathbf{Y}$ and the matrix $\mathbf{\Phi}$, respectively, as shown in equation (6) in the main text and equation (3) for SIS and CP dynamics, respectively.

For another base string $S_{-i}(\hat{t}_\beta)$, we can find t_ν associated with the strings that belong to the base string $S_{-i}(\hat{t}_\beta)$ in the same manner. Based on $\hat{t}_\beta$ and t_ν, we can obtain the two average quantities. Repeating this process for all the base strings gives rise to the reconstruction form $\mathbf{Y} = \mathbf{\Phi} \cdot \mathbf{X}$. Figure 1 in the main text is a schematic illustration of the process. It is noteworthy that different base strings may share some strings in their contribution to the two average quantities in the sense that a string according to the threshold Δ may be subject to more than one base string. This will not have any negative impact on full reconstruction. More importantly, the sharing of strings by different base strings considerably reduces the minimum requirement of the length of time series and makes all available data sufficiently used, accounting for the slow increase of the minimum length $t_{\min}$ of time series with the network size, as shown in Supplementary Fig. 4.

Supplementary Note 4: Setting cut-off values

Identifying links. After the neighboring vector $\mathbf{X}$ of each node has been reconstructed, we need to distinguish the existent links from null connections. This can be done when there is an explicit gap between the values of the reconstructed elements associated with existent links and those with null links so that a cut-off can be set in the gap. The elements with values larger than the cut-off value are regarded as corresponding to actual links, while those with values smaller than the cut-off value correspond to zeros in the adjacency matrix. However, ambiguity may arise due to the dispersion in the element values. It is thus necessary to determine the cut-off value reliably. In the ideal situation, the element values in $\mathbf{X}$ exhibit a bimodal type of distribution, with one peak (centered at some positive value) corresponding to actual links and another (centered about zero) to null connections. However, due to the fluctuations in the distribution, the two peaks may overlap, making it difficult to set a cut-off. Our idea is to use coarse-graining to suppress noise but preserve the useful information. As shown in Supplementary Fig. 1(a), we calculate the local accumulation of the element values in the entire distribution region. The size of each bin can be adjusted according to the width of the peak centered at zero, with the constraint that the bin size should not be larger than the width of this peak. Empirically, 5 to 10 bins are effective to reduce noise while retaining the information about the link structure. Implementing such a coarse-grained process can in general make the "valley" between the two peaks distinct. The coarse-graining works even when there was no clear gap between the two peaks. In this case, we choose the lowest point in the valley between the two peaks to be the cut-off value.

In general, the predicted values of null connections are sufficiently close to zero, enabling us to set a cut-off for different networks, as shown in Supplementary Fig. 1(b). We see that the four different types of networks share the same cut-off value of about 0.13. This cut-off value offers a reasonably way to separate the existent links from null connections, offering a robust procedure to identify actual links in the reconstructed vector $\mathbf{X}$.

Locating hidden source. Our method to ascertain and locate hidden source is based on a statistical analysis. In particular, from different segments of the time series, we calculate the structural variance defined in equation (10) in the main text. For any neighboring node of the hidden source, the value of σ_i is usually much larger than those of nodes that are not in the immediate neighborhood of the hidden source. The σ_i values can thus be used to reliably identify the neighboring nodes of the hidden source. Similar to the task of distinguishing actual links from null connections, we generate the statistical distribution of σ_i and identify those nodes with abnormally large σ_i values, which can be accomplished by using the coarse-grained method, as shown in Supplementary Fig. 1c. After accumulating the bins, a well-defined gap between the small and rare large values of σ_i emerges. A cut-off can then be set within the largest gap, where the nodes associated with the large σ_i values are identified as the immediate neighbors of the hidden source. The trade-off measures of locating the hidden source in Table I in the main text are result of this statistical analysis.

An alternative approach to identifying the immediate neighbors of the hidden source is to define the

following quantity:

$$\sigma_i^* = \frac{1}{N}\sum_{j=1}^{N}\sqrt{\frac{1}{g}\sum_{k=1}^{g}\left(x_{ij}^{(k)} - \langle x_{ij}\rangle\right)^2}, \tag{5}$$

where j denotes the column, $x_{ij}^{(k)}$ represents the element values in the vector $\mathbf{X}_i$ of node i inferred from the kth group of the data, $\langle x_{ij}\rangle = (1/g)\sum_{k=1}^{g} x_{ij}^k$ is the mean value of x_{ij}, and g is the number of data segments. As shown in Supplementary Fig. 1(d), there is a clear gap between the σ_i^* values associated with the neighbors and those with nodes not in the immediate neighborhood. Analogous to the method based on σ_i, different networks share approximately the same cut-off value of about $\sigma_i^* = 0.9$.

Supplementary Note 5: Additional results of network reconstruction

Reconstruction accuracy depending on $n_{\hat{t}}$ for a variety of networks. Supplementary Figs. 2 and 3 show the success rates of reconstructing a number of model and real networks as a function of the fraction $n_{\hat{t}}$ of base strings. For all networks studied, high recovery accuracy in a wide range of the values of $n_{\hat{t}}$ is achieved. However, for $n_{\hat{t}}$ close to unity, there is a slight decrease in the success rates. This can be attributed to the intrinsic characteristics of the compressed-sensing algorithm. Recall that compressed sensing is developed for solving under-determined systems with many more unknowns than the number of equations [12]. In order to assure a unique solution of the under-determined system, the constraint of sparsity must be imposed. Mathematically, if under such a constraint a unique solution indeed exists, the compressed-sensing framework enables recovery of the solution. For $n_{\hat{t}} = 1$, the matrix $\mathbf{\Phi}$ becomes a square matrix, indicating that the system is no longer under-determined. Although the compressed-sensing algorithm is not restricted to under-determined systems, our results demonstrate that it performs better for under-determined systems than for $n_{\hat{t}} = 1$, as reflected by the decrease in the success rates when $n_{\hat{t}}$ approaches to unity. However, since our method would typically be applied with limited data, the reconstruction accuracy can be still assured by using $n_{\hat{t}} < 1$ (so that $n_{\hat{t}} = 1$ is naturally avoided) in the process of setting up the compressed-sensing equation $\mathbf{Y} = \mathbf{\Phi} \cdot \mathbf{X}$.

Minimum length of time series. Figure 4 shows the minimum absolute length $t_{\min}$ of time series required for achieving 95% success rate of the geometric averages of SREL and SRNC versus the network size N for SIS and CP dynamics on NW and BA networks. It can be seen that $t_{\min}$ increases slightly as N is increased and the slope in a logarithmic plot is much less than 1, indicating low data requirement even for large networks. This can be explained by the fact that different base strings may share some common strings in the reconstruction. The relative minimum length $n_t^{\min}$ resulting from $t_{\min}/N$ is shown in Figs. 3(c,d) in the main text. One can see that $n_t^{\min}$ is considerably reduced as N is increased. In general, when a sufficient number of base strings are available so that other strings can be chosen using the threshold parameters in time series, our reconstruction method can offer quite high accuracy, nearly regardless of the network size.

Success rates affected by average node degree. Supplementary Fig. 5 shows the success rates as a function of the average node degree $\langle k \rangle$ for SIS and CP dynamics on ER, NW and BA networks. The success rates decrease from 1 slightly as $\langle k \rangle$ is increased to a large value, e.g., $\langle k \rangle = 18$. The slight decrease is due to two factors.

First, for high values of $\langle k \rangle$, the vector $\mathbf{X}$ is less sparse, leading to small prediction errors under the compressed-sensing paradigm. Second, for a node with a large number of neighbors, the probability of its being infected at each time is large, insofar as a fraction of its neighbors are in the infected state. In this case, it is difficult to ascertain which neighbor passes the disease to the node. For example, in the extreme case where a node remains infected all the time, it is practically impossible to determine from which of its neighbors the infection comes. The combined effect of the two factors accounts for the small decrease in the success rates with the increase of $\langle k \rangle$.

In scale-free networks, a small fraction of hub nodes are usually more difficult to be reconstructed

because of the many neighbors that each hub possesses, leading to slightly lower success rates than those associated with ER and NW networks of similar average degree (e.g., $\langle k \rangle < 14$). However, for larger values of the average degree, scale-free networks perform better than the other two types of networks. This can be understood, as follows. In a scale-free network, the heterogeneous degree distribution indicates that most nodes have small degrees. In contrast, in the ER and NW networks, the degree distributions are approximately homogeneous. Thus, for the same value of $\langle k \rangle$, the degrees of most nodes in scale-free networks are less than those in the ER and NW networks, resulting in slightly larger success rates when reconstructing scale-free networks.

In fact, in the presence of high-degree nodes, we can provide a heuristic method to improve the success rates by taking advantage of the "conflicts" arising when matching the neighborhoods of different nodes, since the neighborhood of each node is recovered independently of others. The errors in the reconstruction of the whole network will induce certain inconsistency in predicting the presence/absence of links among nodes. Due to the relatively large reconstruction errors associated with higher-degree nodes, it is reasonable to accept the predicted links centered at smaller-degree nodes rather than that centered at high-degree nodes whenever such a conflict arises. This way the errors associated with high-degree nodes can be considerably reduced. This method can further improve the reconstruction accuracy of scale-free networks or in any networks where a small set of high-degree nodes exist.

Reconstruction of vector $\mathbf{X}$ in the presence of partially unobservable nodes. Supplementary Fig. 6 shows the element values in the reconstructed vector $\mathbf{X}$ when there are missing time series. In principle, the local connection topology of nodes without time series are impossible to be reconstructed. We thus remove these nodes together with their links from the network and reconstruct the rest of network. Interestingly, we find that even when half of nodes' time series are missing, i.e., $n_{\rm m} = 0.5$, there is still a clear gap between the predicted values of actual links and null connections. This phenomenon suggests that missing time series only affect the width of the gap, but otherwise has little influence on the reconstruction. Indeed, in Figs. 4(c,d) in the main text, the success rates remain close to 1 as $n_{\rm m}$ is increased.

We stress that the externally inaccessible nodes are somewhat similar to the hidden source in the sense that their time series are not available. The difference between the hidden source and the unobservable nodes lies in the fact that the source is always infected with relatively higher probability to pass the disease to its neighbors. The absence of connection from the source when establishing the reconstruction form of its immediate neighbors results in the large structural variance that becomes the fingerprint of the immediate neighbors. In contrast, the effect of unobservable nodes is relatively minor as compared to that of the hidden source, so their connections to the other observable nodes are difficult to be inferred. Consequently, in the presence of externally inaccessible nodes, the structural variance still allows us to exclusively locate the hidden source.

Robustness with respect to fluctuation in infection rates. We have also investigated robustness of our reconstruction method with respect to errors in specification of the underlying diffusion process. This is of practical importance, since it is not possible to describe any realistic dynamical process by mathematical model with perfect accuracy. To be concrete, we assume that, in the spreading process,

the infection rate λ of each node exhibits random fluctuations in a range with probability p_λ at each time step. In this case, ideally the standard SIS and CP dynamics are no longer applicable. However, since such fluctuations are not known a priori, we reconstruct the dynamical process as if the infection rates were constant. This way the effect of model errors can be assessed. Supplementary Fig. 7 shows the success rate of network reconstruction as a function of p_λ. We see that quite satisfactory reconstruction of the network structure can still be achieved, regardless of the value of p_λ. This result holds for both SIS and CP processes on small-world and scale-free networks, suggesting that our reconstruction method is applicable to real propagation processes in spite of model errors.

Supplementary Note 6: Inferring infection and recovery rates from time series

Inferring infection rates. After all links have been successfully predicted, the infection rates λ_i can be recovered by equation (17) for SIS dynamics and equation (18) for CP dynamics in the main text. Supplementary Fig. 8 shows the standard box-plots of differences between the actual and predicted λ_i values for the two types of dynamics coupled with three types of network topologies. We find that for ER and NW networks, the predicted λ_i values of all nodes are quite close to the actual values. In contrast, for BA networks and SIS dynamics, the predicted values of λ_i associated with a small number of high-degree nodes have relatively large errors. This is due to the inaccuracy in the reconstruction of the neighboring nodes of the high-degree nodes. However, the prediction errors can be reduced by replacing links centered at high-degree nodes by the recovered links centered at the neighbors of these nodes. Insofar as the neighborhood of a node is precisely predicted, the value of λ_i associated with the node can be accurately reproduced as well.

Inferring recovery rates. The heterogeneous recovery rates of nodes can be inferred from time series in a more straightforward manner than the infection rates. This is because the recovery rate δ_i of an arbitrary node i is an intrinsic characteristic of the node, which does not depend on the properties of its neighboring nodes. The information about δ_i is contained in the transition from the infected to the susceptible states associated with the probability P^{10}. We thus propose the following strategy to estimate δ_i:

$$\delta_i = P_i^{10} = 1 - P_i^{11} = 1 - \frac{\sum_{j=1}^{n-1} S_i(t_j) S_i(t_j+1)}{\sum_{j=1}^{n-1} S_i(t_j)}, \tag{6}$$

where n is the length of the time series, P^{11} is the probability for the node to remain infected, $\sum_{j=1}^{n-1} S_i(t_j)$ $S_i(t_j+1)$ denotes the number of 11 pairs and $\sum_{j=1}^{n-1} S_i(t_j)$ represents the total number of 10 and 11 pairs in the time series. Equation (6) means that the recovery rate can be assessed by the fraction of 10 pairs in the total number of 10 and 11 pairs in the time series. The predicted recovery rates are shown in Supplementary Fig. 9 for SIS and CP dynamics coupled with three types of networks. The standard box-plots demonstrate that the reproduced recovery rates are in good agreement with the true rates with negligibly small differences for all studied cases. Our method thus enables accurate reconstruction of the natural diversity of nodes in terms of the heterogeneous infection rates, the recovery rates, and the node degrees.

It is noteworthy that, when the recovery rates $\delta_i(t)$ of individuals are time-dependent, our method can still be used to reconstruct the network structure and the quantitative measures of the individual diversity. This is because our reconstruction framework does not require any information about the recovery rates. In fact, all useful information for uncovering the network structure and the infection rates is contained in the 00 and 01 pairs rather than in the 10 or 11 pairs, where the latter are used to estimate the recovery rates. The robustness of our reconstruction framework against stochastic fluctuations in the recovery rates of individuals renders applicable our framework to realistic stochastic systems.

Supplementary Note 7: Monte Carlo simulations for SIS and CP

Numerical simulation of SIS. Initially, a fraction of nodes are set to be infected (state 1), where a node in this group can be either the external source or a node within the network. Nodal states are updated in parallel. For a susceptible node i of degree k_i, at time t, if m neighbors are infected, the probability that it will be infected by each of the m neighbors is λ_i. For each infected neighbor, a random number is drawn and checked if it is smaller than λ_i. If yes, node i will become infected at the next time step; otherwise, nothing happens and the process is repeated for the next infected neighbor until i becomes infected or all infected neighbors have been tested. At the same time t, an infected node j has the rate δ_j to recover to the susceptible state at the next time step.

Numerical simulation of CP. The initial configuration and state updating rules are the same as those for SIS. At time t, for a susceptible node i, one of its neighbors is randomly chosen. If the chosen neighbor is infected, i will become infected at time $t+1$ with the probability λ_i. Computationally, a random number is drawn to compare with λ_i: if the number is smaller than λ_i, i will be infected at $t+1$; otherwise, i will remain in the susceptible state. The recovery rule is the same as that for SIS. A Monte Carlo time t is referred to the situation where all the states at $t+1$ have been updated according to their states at t.

We have carried out a large number of simulations to validate our reconstruction framework. We find that our approach is robust and broadly applicable to various combinations of the network structure and spreading-dynamical processes. The results are generally independent of the parameter setting of the network model and the initial configuration of the diffusion dynamics.

Supplementary Note 8: The approximation $\langle \ln[1-P_i^{01}(t)]\rangle \simeq \ln[1-\langle P_i^{01}(t)\rangle]$

For SIS dynamics, to derive equation (15) in the main text, we have used the following approximation:

$$\langle \ln[1-P_i^{01}(t)]\rangle \simeq \ln[1-\langle P_i^{01}(t)\rangle], \tag{7}$$

where $P_i^{01}(t)$ is the infection probability of node i at time t. Equation (7) is valid if $P_i^{01}(t)$ is relatively small. For simplicity, we denote $P_i^{01}(t)$ by $\bar{P}$ and the ensemble average involves $\bar{P}_\nu$ for $\nu = 1, 2, \cdots, l$. We can write

$$\begin{aligned} \langle \ln[1-P_i^{01}(t)]\rangle &= \frac{\ln(1-\bar{P}_1)+\ln(1-\bar{P}_2)+\cdots+\ln(1-\bar{P}_l)}{l} \\ &= \ln\left[(1-\bar{P}_1)^{\frac{1}{l}}(1-\bar{P}_2)^{\frac{1}{l}}\cdots(1-\bar{P}_l)^{\frac{1}{l}}\right]. \end{aligned} \tag{8}$$

Because of $0 < \bar{P}_\nu < 1$, according to the binomial expression, we have

$$(1-\bar{P}_\nu)^{\frac{1}{l}} = \sum_{k=0}^{\infty} \begin{pmatrix} \frac{1}{l} \\ k \end{pmatrix} \left(-\bar{P}_\nu\right)^k = 1 - \frac{1}{l}\bar{P}_\nu + \frac{1-l}{2l^2}\bar{P}_\nu^2 + \mathcal{O}^3. \tag{9}$$

If $\bar{P}_\nu$ is relatively small, keeping only the zero-order and first-order terms suffices, yielding

$$(1-\bar{P}_\nu)^{\frac{1}{l}} \simeq 1 - \frac{\bar{P}_\nu}{l}. \tag{10}$$

Inserting equation (10) into equation (8), we have

$$\begin{aligned} \ln\left[(1-\bar{P}_1)^{\frac{1}{l}}(1-\bar{P}_2)^{\frac{1}{l}}\cdots(1-\bar{P}_l)^{\frac{1}{l}}\right] &\simeq \ln\left[\left(1-\frac{\bar{P}_1}{l}\right)\left(1-\frac{\bar{P}_2}{l}\right)\cdots\left(1-\frac{\bar{P}_l}{l}\right)\right] \\ &\simeq \ln\left[1-\frac{\bar{P}_1+\bar{P}_2+\cdots+\bar{P}_l}{l}+\mathcal{O}^2\right] \\ &\simeq \ln\left[1-\langle\bar{P}_\nu\rangle+\mathcal{O}^2\right]. \end{aligned} \tag{11}$$

Since $\bar{P}_\nu < 1$, $\bar{P}_\nu/l \ll 1$, we can omit $\mathcal{O}^2$ terms. As a result, equation (11) indicates that approximation [equation (7)] is justified for relatively small values of the infection probability.

We have also numerically tested the approximation, as shown in Supplementary Fig. 10. We can see that when the infection rate λ_i is randomly distributed in the range (0.2,0.4), both sides of equation (7) are quite close to each other with small deviation from the diagonal line. Moreover, even if the fitted line deviates from the diagonal line, the linear correlation can still assure the success of reconstruction based on equation (15) in the main text. The difference between the slope of the fitted lines and unity will become a small coefficient added to the neighboring vector $\mathbf{X}$, which will slightly affect the element values in $\mathbf{X}$, affecting only the width of the gap between actual links and null connections, but not the cut-off and identification of links.

Supplementary Note 9: Dependence of success rates on threshold parameters

In our reconstruction framework, there are two thresholds, Δ and Θ, which are used in the calculation of the normalized Hamming distance. Here, we discuss the possible region in the parameter plane spanned by the two thresholds, which ensures full reconstruction of the network. Each threshold value is bounded in the unit interval, as stipulated by the definition of the normalized Hamming distance. Specifically, Δ is employed to approximate the law of large numbers because of the difficulty to find two strings with absolute zero normalized Hamming distance. With a string $S_{-i}(\hat{t}_\alpha)$ chosen to be a base, we select suitable strings through the threshold Δ:

$$H[S_{-i}(\hat{t}_\alpha), S_{-i}(t)] < \Delta, \forall t \in T(\hat{t}_\alpha). \tag{12}$$

We can then use the states $S_i(t)$ of node i at time t belonging to the available strings to estimate the infection probability of i at t_α by

$$\langle S_i(\hat{t}_\alpha + 1)\rangle \simeq P_i^{01}(\hat{t}_\alpha). \tag{13}$$

The threshold Θ is used to provide linearly independent equations containing structural information to be revealed by the compressed sensing algorithm. In general, an equation can be found based on a base string, e.g., $S_{-i}(\hat{t}_\alpha)$. We choose a series of base strings to construct a set T_{base}, in which each pair of vectors satisfy

$$H[S_{-i}(\hat{t}_\beta), S_{-i}(\hat{t}_\alpha)] > \Theta, \quad \forall \hat{t}_\alpha, \hat{t}_\beta \in T_{base},$$

where $\hat{t}_\alpha$ and $\hat{t}_\beta$ stand for the time of the two base vectors in the time series. The number of nontrivial equations is determined by the number of base strings.

The dependence of the success rates on the values of the thresholds Δ and Θ is shown in Supplementary Figs. 11 and 12 for SIS and CP dynamics, respectively. If and only if both SREL and SRNC approach unity, the network is regarded as having been exactly reconstructed. In Supplementary Figs. 11 and 12, the smaller value of SREL and SRNC is defined to be the success rate. We see that for both SIS and CP dynamics coupled with four types of networks, there is a broad region in the (Δ-Θ) parameter plane that warrants full reconstruction with success rate close to 100%, indicating that our framework is reliable.

The broad region leading to full reconstruction in the parameter plane implies that our framework is robust against measurement noise embedded in the binary time series of nodal states. Suppose that a string belongs to a base string in the sense that the normalized Hamming distance between them is less than the threshold Δ. In the presence of noise, some states of nodes in the strings are reversed, resulting in the change of the normalized Hamming distance that may become larger than the threshold Δ. However, due to the broad region of the applicable values of Δ, the change of the normalized Hamming distance is tolerable and the string is likely to still belong to the base string and can be used to estimate the infection probability $P^{01}(t)$, so the selection of strings with respect to a base string based on the normalized Hamming distance is robust against noise. Analogously, identifying different base strings in terms of the threshold Θ is insensitive to noise due to the broad region of Θ. The other factor by which noise may affect the reconstruction accuracy is the process of averaging over node states for obtaining $\langle S_i(\hat{t}_\alpha + 1)\rangle$

and $\langle S_i(\hat{t})\rangle$. Since the error in measuring node states induced by noise will be naturally reduced by the average, there is resistance to noise. In general, noise can be tolerated in all steps required to construct the vector $\mathbf{Y}$ and matrix $\mathbf{\Phi}$, ensuring robustness of our reconstruction framework against noise in time series.

Supplementary Note 10: Network models and data sets

ER network model [15]. In the ER random-network model, N isolated nodes are given initially. Each pair of nodes are connected with probability p ($0 < p < 1$). The average node degree $\langle k \rangle$ is approximately Np. The node degrees obey the Poisson distribution.

WS network model [16]. Initially a ring-like structure of N nodes is constructed, where every node has the same number of local connections. Each link is then randomly rewired with probability p. For p close neither to zero nor to one, the network exhibits the small-world effect, i.e., highly clustered structures but with short average distance. In our work, $p = 0.2$ is used for obtaining all the results on the small-world WS networks.

NW network model [17]. The NW model is capable of generating small-world networks while keeping the network as a single component. A ring-like structure of N nodes is generated initially, where the number of local connections of every node is the same as that of any other node. Random links are then added into the network. The average degree of the network is determined by the initial local connections and the number of shortcuts added. The initial ring is of coordinate number 1 and for each node, a link from the selected node is added to another randomly selected node. In our paper, the average node degree is $\langle k \rangle = 4$ for NW networks.

BA network model [18]. The BA model generates scale-free networks with power-law degree distribution. Initially, a core consisting of a small number of nodes with random connections is constructed. At each time step, a new node with m links is added to the existent network according the preferential-attachment mechanism: the probability of a new node connecting to an existent node is proportional to its degree. Nodes are continuously added to the network until the pre-specified network size N is reached. The average node degree is $\langle k \rangle = 2m$.

Real networks. The details of the real social networks studied in this paper are presented in Supplementary Table 1, which includes the number N of nodes, the number L of links, and description of the networks.

Supplementary References

1. Zachary, W. W. An information flow model for conflict and fission in small groups. *J. Anthropol. Res.* 452–473 (1977).

2. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.* **99**, 7821–7826 (2002).

3. Milo, R. *et al.* Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).

4. van Duijn, M. A., Zeggelink, E. P., Huisman, M., Stokman, F. N. & Wasseur, F. W. Evolution of sociology freshmen into a friendship network. *J. Math. Sociol.* **27**, 153–191 (2003).

5. Lusseau, D. *et al.* The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations. *Behav. Ecol. Sociobiol.* **54**, 396–405 (2003).

6. http://www.orgnet.com/divided.html.

7. Newman, M. E. J. Finding community structure in networks using the eigenvectors of matrices. *Phys. Rev. E* **74**, 036104 (2006).

8. Candès, E. J., Romberg, J. & Tao, T. Robust uncertainty principles: Exact signal reconstruction from highly incomplete frequency information. *IEEE Trans. Inf. Theo.* **52**, 489–509 (2006).

9. Baraniuk, R. G. Compressive sensing [lecture notes]. *Sig. Proc. Mag. IEEE* **24**, 118–121 (2007).

10. Candès, E. J. & Wakin, M. B. An introduction to compressive sampling. *Sig. Proc. Mag. IEEE* **25**, 21–30 (2008).

11. Romberg, J. Imaging via compressive sampling. *Sig. Proc. Mag. IEEE* **25**, 14–20 (2008).

12. Candès, E. J. Compressive sampling. In *Proceedings of the International Congress of Mathematicians: Madrid, August 22-30, 2006: invited lectures*, 1433–1452 (2006).

13. Candes, E. & Romberg, J. l1-magic: Recovery of sparse signals via convex programming. *URL: www. acm. caltech. edu/l1magic/downloads/l1magic. pdf* **4** (2005).

14. l1-Magic - ECE Users Pages. *URL: http://users.ece.gatech.edu/ justin/l1magic/.*

15. Erdës, P. & Rényi, A. On the evolution of random graphs. *Publ. Math. Inst. Hungar. Acad. Sci* **5**, 17–61 (1960).

16. Watts, D. J. & Strogatz, S. H. Collective dynamics of small-worldnetworks. *Nature* **393**, 440–442 (1998).

17. Newman, M. E. & Watts, D. J. Renormalization group analysis of the small-world network model. *Phys. Lett. A* **263**, 341–346 (1999).

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