Supplementary Material for

Robust Reconstruction of Complex Networks from Sparse Data

Contents

1 Performance assessment

To quantify the performance of our reconstruction method, we introduce two standard measurement indices, the area under the receiver operating characteristic curve (AUROC) and the area under the precision-recall curve (AUPR) [S1]. True positive rate (TPR), false positive rate (FPR), Precision and Recall that are used to calculate AUROC and AUPR are defined as follows:

$$\mathrm{TPR}(l) = \frac{\mathrm{TP}(l)}{P}, \tag{S1}$$

where l is the cutoff in the edge list, $\mathrm{TP}(l)$ is the number of true positives in the top l predictions in the edge list, and P is the number of positives in the gold standard.

$$\mathrm{FPR}(l) = \frac{\mathrm{FP}(l)}{Q}, \tag{S2}$$

where $\mathrm{FP}(l)$ is the number of false positive in the top l predictions in the edge list, and Q is the number of negatives in the gold standard.

$$\mathrm{Precision}(l) = \frac{\mathrm{TP}(l)}{\mathrm{TP}(l) + \mathrm{FP}(l)} = \frac{\mathrm{TP}(l)}{l}, \tag{S3}$$

$$\mathrm{Recall}(l) = \frac{\mathrm{TP}(l)}{P}, \tag{S4}$$

where $\mathrm{Recall}(l)$, which is called sensitivity, is equivalent to $\mathrm{TPR}(l)$.

2 Detailed results in addition to Table I and II in the main text

Table I in the main text shows the minimum data amount that ensure at least 95% AUROC and AUPR simultaneously for different cases. In Supplementary Materials, we provide more details of AUROC and AUPR as a function of data amount for all the cases presented in Table I in the main text. Supplementary Fig. S2- S4 show for network size $N = 100$ and average node degree $\langle k \rangle = 6$, AUROC and AUPR as a function of data amount for different variance σ^2 of Gaussian white noise $\mathcal{N}(0, \sigma^2)$ embedded in the time series for obtaining vector $\mathbf{Y}$ in the reconstruction form $\mathbf{Y} = \Phi\mathbf{X}$. For three types dynamical processes, evolutionary ultimatum games, transportation and communications, full reconstruction of network structure can be achieved from sufficient data in three types model networks. In the absence of noise or for small noise variance, say $\sigma = 0.05$ (Supplementary Fig. S2 and Fig. S3), full reconstruction can be assured by small amounts of data relative to the network size N. For large noise variance, say, $\sigma = 0.3$ (Supplementary Fig. S4), full reconstruction can still be achieved based on relatively large amounts of data for different networks, manifesting the strong robustness of our method against noise in time series.

We have also tested the robustness of our reconstruction against the existence of externally inaccessible nodes. We assume that a fraction of randomly selected nodes cannot be measured and their time series are missing. Supplementary Fig. S5 and Fig. S6 show AUROC and AUPR as a function of data amount for $n_{\mathrm{m}} = 5\%$ and $n_{\mathrm{m}} = 30\%$ fraction of inaccessible nodes, respectively. We see that even for $n_{\mathrm{m}} = 30\%$, we can still accurately reconstruct connections among available nodes from time series of three dynamical processes, indicating strong robustness of our method against the missing of partial nodes. Furthermore, based on the reconstructed network, we can identify who has connections with unobservable nodes in evolutionary ultimatum games. This can be accomplished by comparing actual $\mathbf{Y}$ with reconstructed $\mathbf{Y}'$ from time series. To be concrete, we can use the time series of players' strategies and reconstructed network to calculate payoffs of a player in different rounds. The calculated $\mathbf{Y}'$ will differ from the actual $\mathbf{Y}$ if the reconstructed node connects to some of the inaccessible nodes, which provides a clue for inferring direct neighbors of missing nodes.

Supplementary Fig. S7 and Fig. S8 show AUROC and AUPR as a function of data amount for average node degree $\langle k \rangle = 12$ and $\langle k \rangle = 18$, respectively, with $N = 100$. The results demonstrate that for large values of $\langle k \rangle$, our method can guarantee complete identification of all links for different combinations of network structures and dynamical processes. Supplementary Fig. S9 and Fig. S10 show for $\langle k \rangle = 6$, AUROC and AUPR as a function of data amount for the network size $N = 500$ and 1000, respectively. Supplementary Fig. S11 shows the reconstruction performance of our method applied to several empirical networks. We see that for larger network sizes, our method requires relatively less data to ascertain all links. This is due to the fact that the neighboring vector becomes sparser in larger networks. The L_1 norm in the lasso ensures the requirement of less data for reconstructing a sparser vector, allowing us to achieve full reconstruction by using less data for larger networks.

3 Data of empirical networks

In the main text, we have used several empirical networks to test the performance of our reconstruction method. Here we provide more details of these empirical networks, as displayed in Supplementary Table S1.

Supplementary Table S1: Summary of the empirical networks used in Table II in the main text. Here N denotes network sizes, L denotes the number of links and $\langle k \rangle$ represents average degree of a network. Network data can be downloaded from relevant references.

Networks	N	L	$\langle k \rangle$	Class	Description
Karate [S2]	34	78	4.6	Social network	Network of friendship in a karate club
Dolphins [S3]	62	159	5.1	Social network	Frequent associations between 62 dolphins
Netscience [S4]	1589	2742	3.5	Social network	Coauthorship network of scientists working on network
IEEE 39 BUS [S5]	39	46	2.4	Power network	10-machine New-England Power System network
IEEE 118 BUS [S6]	118	177	3.0	Power network	A portion of the American Electric Power System network
IEEE 300 BUS [S7]	300	409	2.7	Power network	Network developed by the IEEE Test Systems Task Force
Football [S8]	115	613	10.7	Social network	Network of American college football game
Jazz [S9]	198	2742	27.7	Social network	Network of jazz musicians
Email [S10]	1133	5451	9.6	Social network	Network of email interchanges

4 Inferring intrinsic individual dynamics in ultimatum games

Insofar as the network structure of evolutionary ultimatum games are ascertained, strategy updating rules of players accompanied with mutation noise can be inferred by a phase diagram. To be concrete, we introduce the phase diagram by defining two variables, as follows:

$$\Delta p_{i,\max} \equiv |p_i(t+1) - p_{\max}(t)| \tag{S5}$$

and

$$\Delta p_{i,i} \equiv |p_i(t+1) - p_i(t)|, \tag{S6}$$

where $|\cdot|$ represents absolute value and $p_{\max}$ is the proposal associated with maximum payoff in the neighborhood of node i. Here $\Delta p_{i,\max}$ represents the absolute difference between $p_i(t+1)$ and $p_{\max}(t)$. If $\Delta p_{i,\max}$ is sufficiently small, e.g., smaller than strategy updating (SU) noise δ (mutation rate), we can confidently infer that player i learned the strategy corresponding to the maximum payoff in i's neighbors at time $t+1$. On the other hand, $\Delta p_{i,i}$ that measures the absolute difference between $p_i(t+1)$ and $p_i(t)$ can be used to ascertain if player i updated his/her strategy at time $t+1$. The fact that $\Delta p_{i,i}$ is smaller than SU noise δ indicates that i didn't update his/her strategy at time $t+1$. Thus the two-parameter phase diagram provides necessary information for revealing how players update their strategies in the evolutionary games. Since we have already had the interactions network amount players, by incorporating the time series of payoffs and strategies, we can identify $p_{\max}(t)$ of each player in each round, allowing us to draw the phase diagram.

As shown in Supplementary Fig. S12, the $\Delta p_{i,\max}$-$\Delta p_{i,i}$ phase diagram is divided into four regions by two boundaries of SU noise. In region (I) ($\Delta p_{i,\max} < \delta$ and $\Delta p_{i,i} > \delta$), data points reflect that individuals update their strategies by imitating those of their neighbors with the highest payoffs. In region (II) ($\Delta p_{i,i} < \delta$ and $\Delta p_{i,\max} > \delta$), data points indicate that players do not vary their strategies in the next round. In region (III) ($\Delta p_{i,\max} < \delta$ and $\Delta p_{i,i} < \delta$), one cannot ascertain the actions of players due to the fact that the difference between $p_i(t+1)$ and $p_{\max}(t)$ is less than the noise δ. However, if data points appear in the diagonal of region (III), the players corresponding to the data points must hold the best strategies in their neighborhoods in the last round. In region (IV) ($\Delta p_{i,\max} > \delta$ and $\Delta p_{i,i} > \delta$), data points indicate players update their strategies by learning from their neighbors without highest payoffs. Despite the existence of SU noise, the phase diagram provides sufficient information to uncover how players update their strategies, regardless of noise. As shown in Supplementary Fig. S12, all data points are presented in region (I), (II) and (III), manifesting that they either learn from the best strategy or keep their own strategies in the last round. The phase diagram can be implemented for each single player, allowing us to identify a mixture of players with different learning inclinations. Although the ground truth of SU noise is unknown, the explicit boundary induced by SU noise in Supplementary Fig. S12 yields the SU noise directly, allowing us to infer intrinsic nodal dynamics associated with mutation noise in strategy updating. If all players learn from best strategies in their neighborhood, there will exist an abrupt transition at the boundaries from region (I) to region (IV) and from region (II) to region (IV). The phase transition can be simply proved to exist in the phase diagram.

5 Supplementary References

[S1] D. Marbach, J. C. Costello and R. Kúffner, *et al*, Nat. Methods, **9**, 796 (2012).

[S2] W. W. Zachary, J. Anthropol. Res. **33**, 452 (1977).

[S3] D. Lusseau, *et al*, Behav. Ecol. Sociobiol **54**, 396 (2003).

[S4] M. E. J. Newman, Phys. Rev. E. **74** 036104 (2006).

[S5] M. A. Pai, *Energy function analysis for power system stability* (Springer, 1989).

[S6] P. M. Mahadev and R. D. Christie, IEEE Trans. Power Syst. **8**, 1084 (1993).

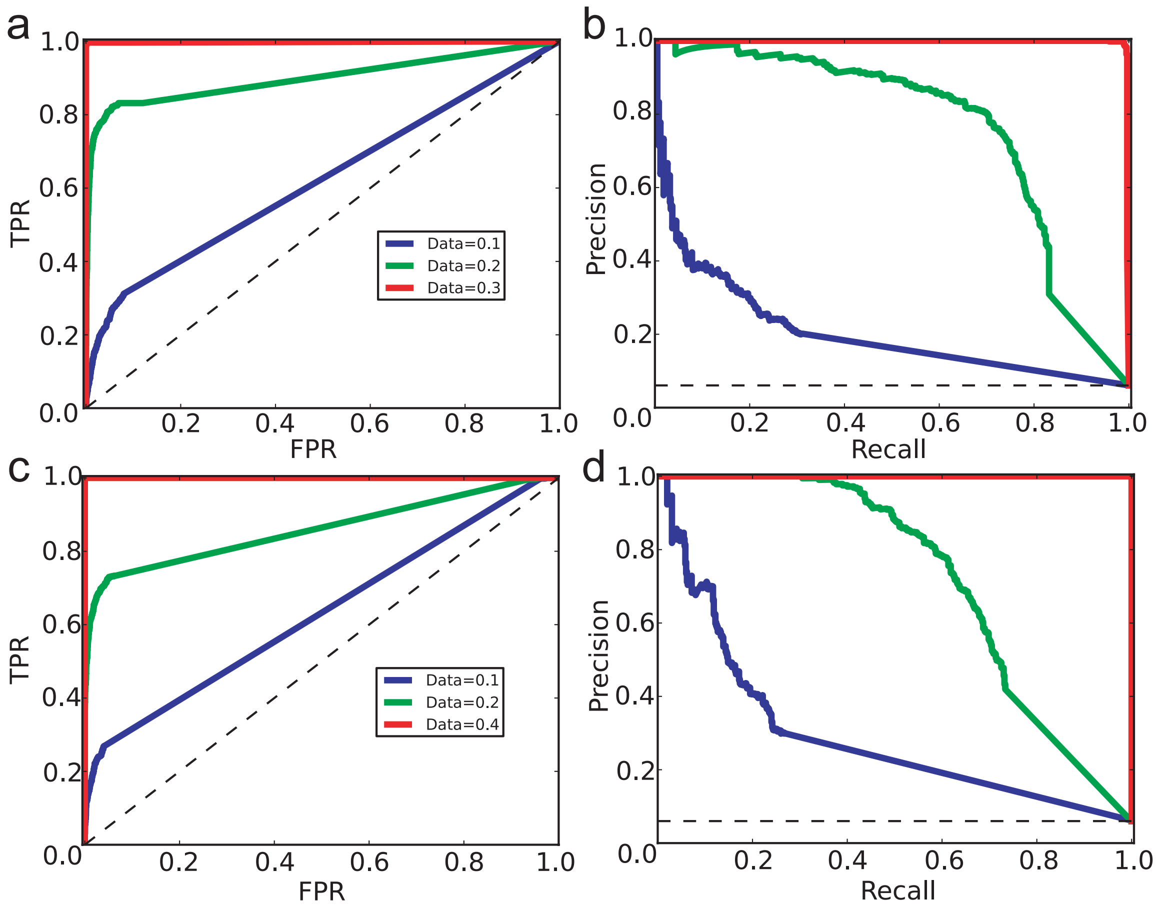
[S7] H. Glatvitsch, F. Alvarado, IEEE Trans. Power Syst. **13**, 1013 (1998).

[S8] M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA **99**, 7821 (2002).

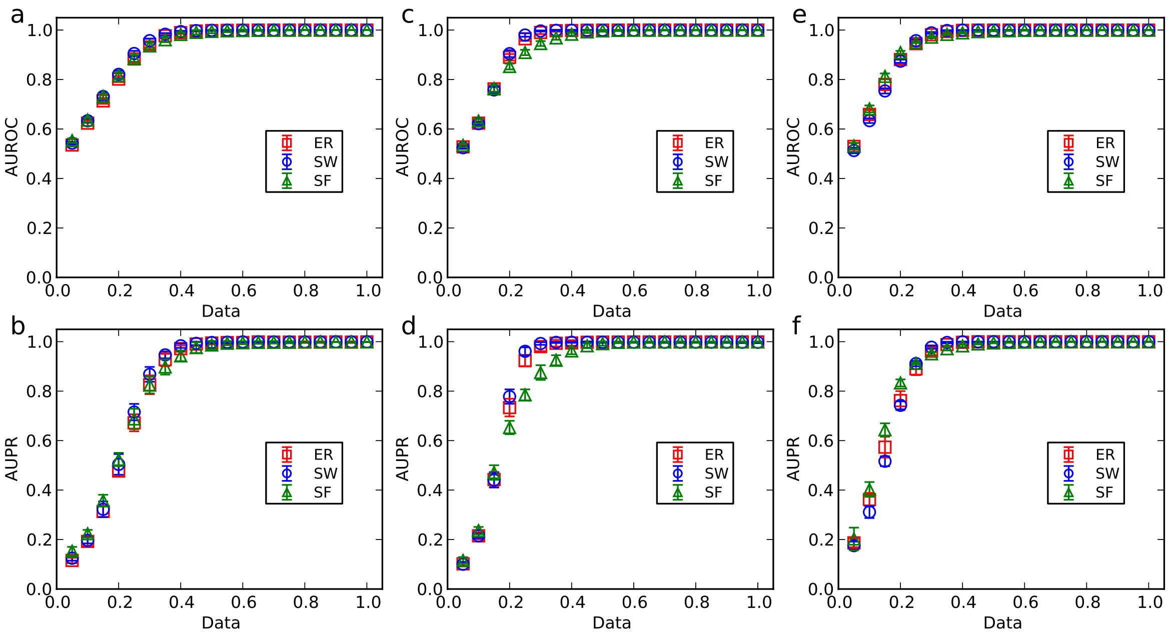
[S9] P. Gleiser and L. Danon, Adv. Complex Syst. **6**, 565 (2003).

[S10] R. Guimera, L. Danon, A. Diaz-Guilera, F. Giralt and A. Arenas, Phys. Rev. E , **68**, 065103, (2003).

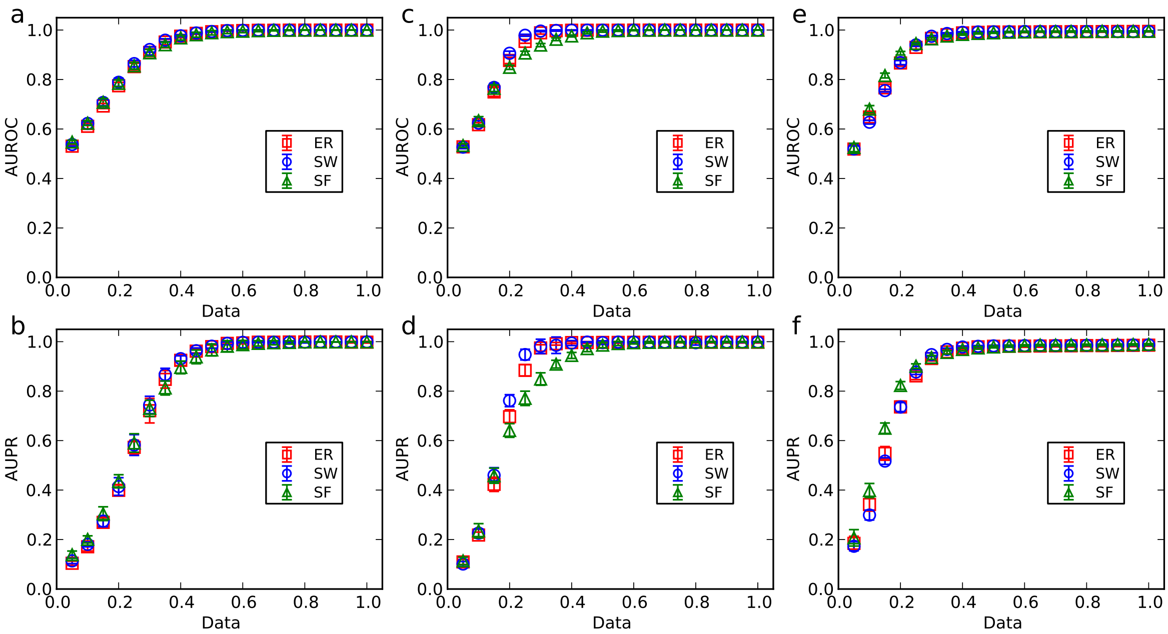
6 Supplementary Figures



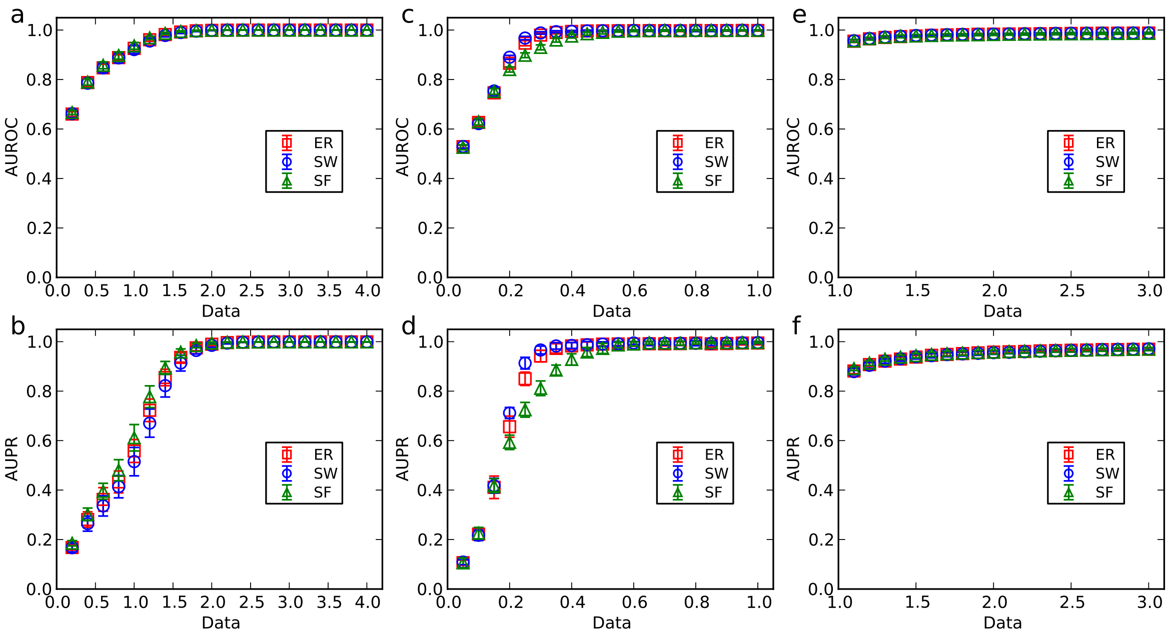
Supplementary Figure S1: (a) AUROC and (b) AUPR of reconstructing resistor networks from different amounts of Data. (c) AUROC and (d) AUPR of reconstructing communication networks from different amounts of Data. Watts-Strogatz small-world networks are used with network size $N = 100$, average degree $\langle k \rangle = 6$, and rewiring probability 0.3. Data is the number of data samplings divided by network size N. The dashed lines represent the results of completely random guesses.



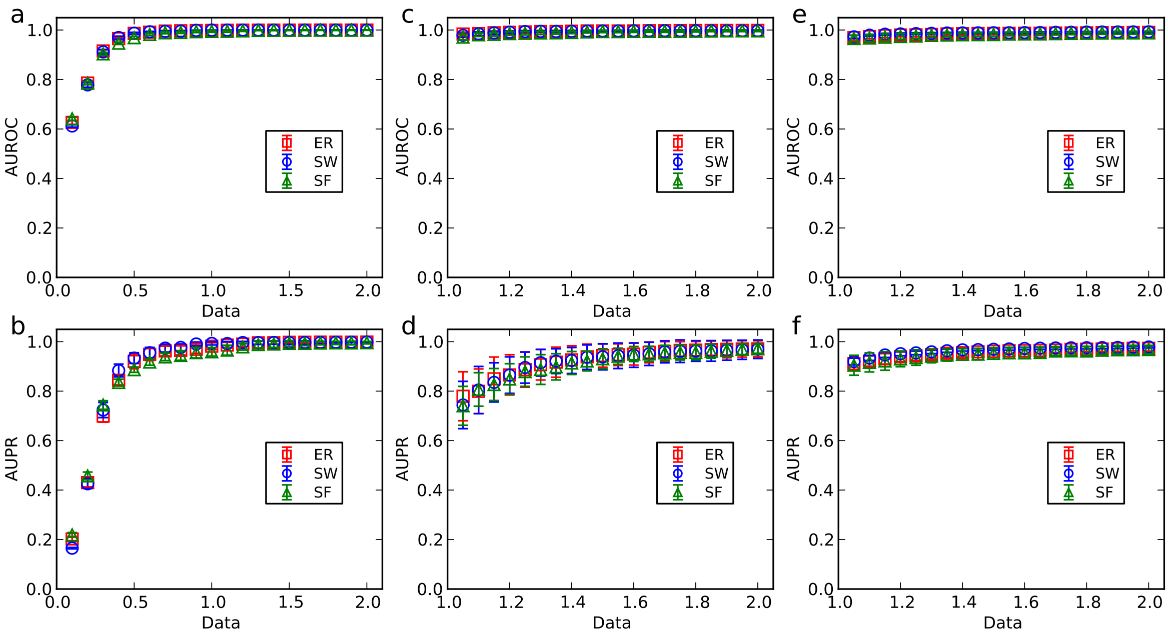
Supplementary Figure S2: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Network size N is 100. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Average degree $\langle k \rangle = 6$ and Rewiring probability of SW networks is 0.3.



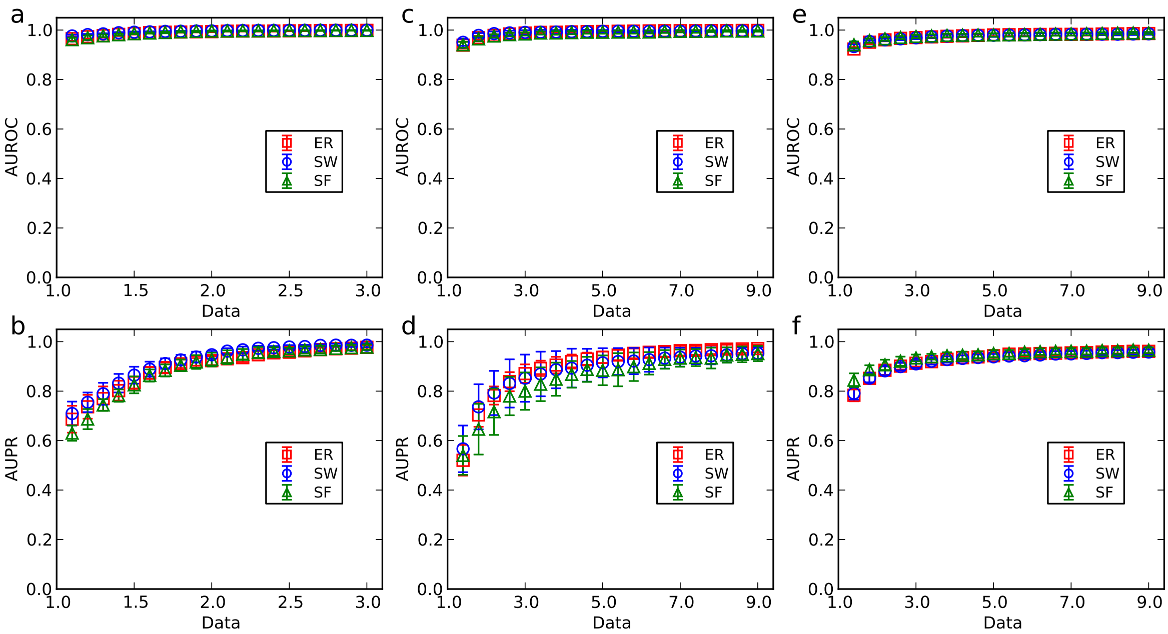
Supplementary Figure S3: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications in the presence of noise. Network size N is 100. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. The average degree $\langle k \rangle = 6$. Rewiring probability of SW networks is 0.3. The distribution of noise is $\mathcal{N}(0, 0.05^2)$.



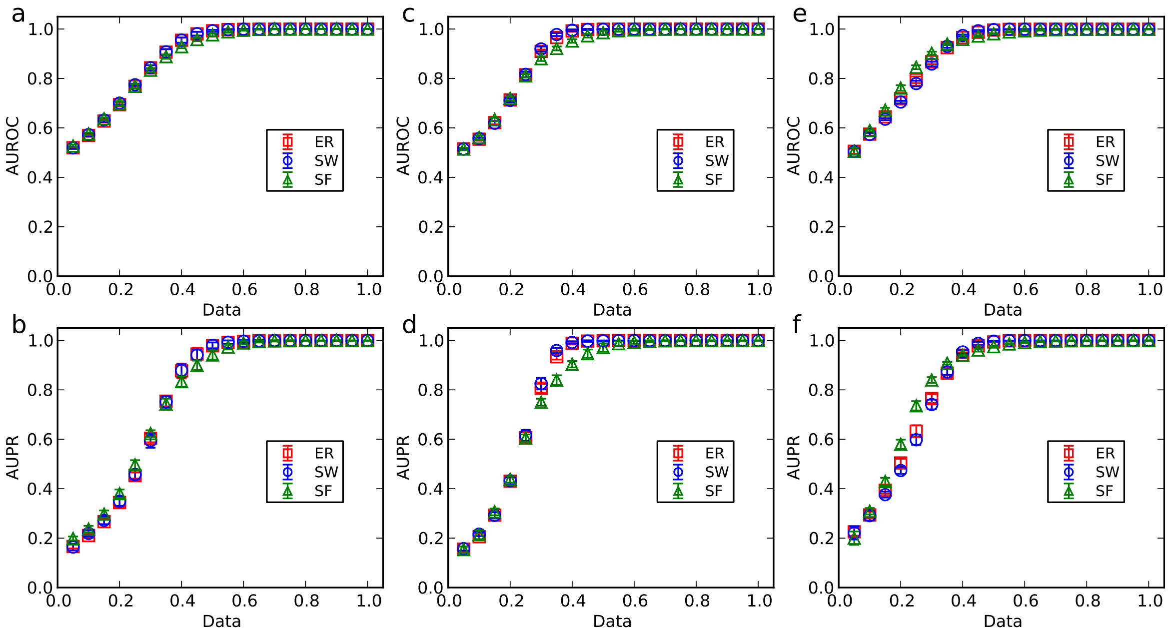
Supplementary Figure S4: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications in the presence of noise. Network size N is 100. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. The average degree $\langle k \rangle = 6$. Rewiring probability of SW networks is 0.3. The distribution of noise is $\mathcal{N}(0, 0.3^2)$.



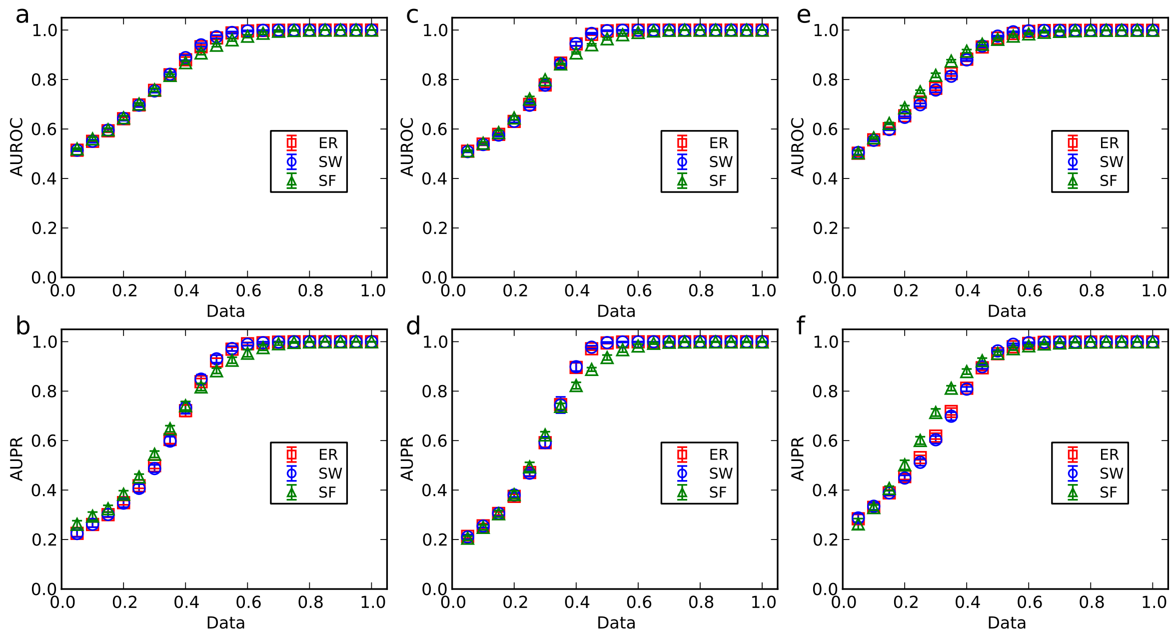
Supplementary Figure S5: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications in the presence of externally inaccessible nodes. The fraction n_{m} of externally inaccessible nodes is 0.05. Network size N is 100. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Average degree $\langle k \rangle = 6$. Rewiring probability of SW networks is 0.3.



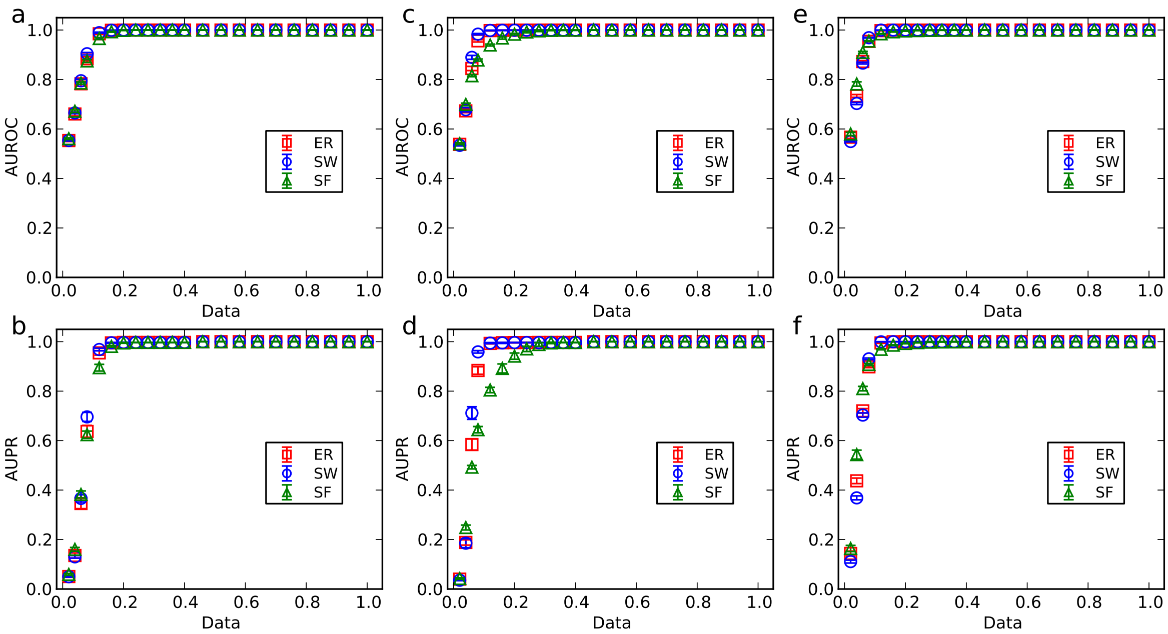
Supplementary Figure S6: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications in the presence of externally inaccessible nodes. The fraction $n_{\rm m}$ of externally inaccessible nodes is 0.3. Network size N is 100. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Average degree $\langle k \rangle = 6$. Rewiring probability of SW networks is 0.3.



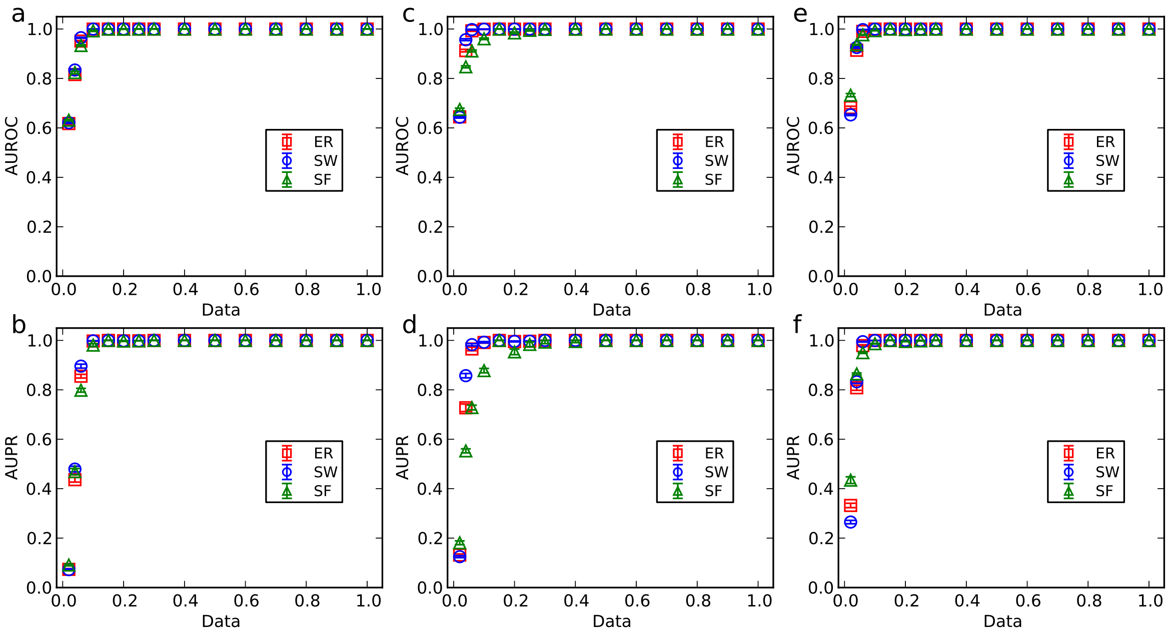
Supplementary Figure S7: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Network size N is 100 and average degree $\langle k \rangle = 12$. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Rewiring probability of SW networks is 0.3.



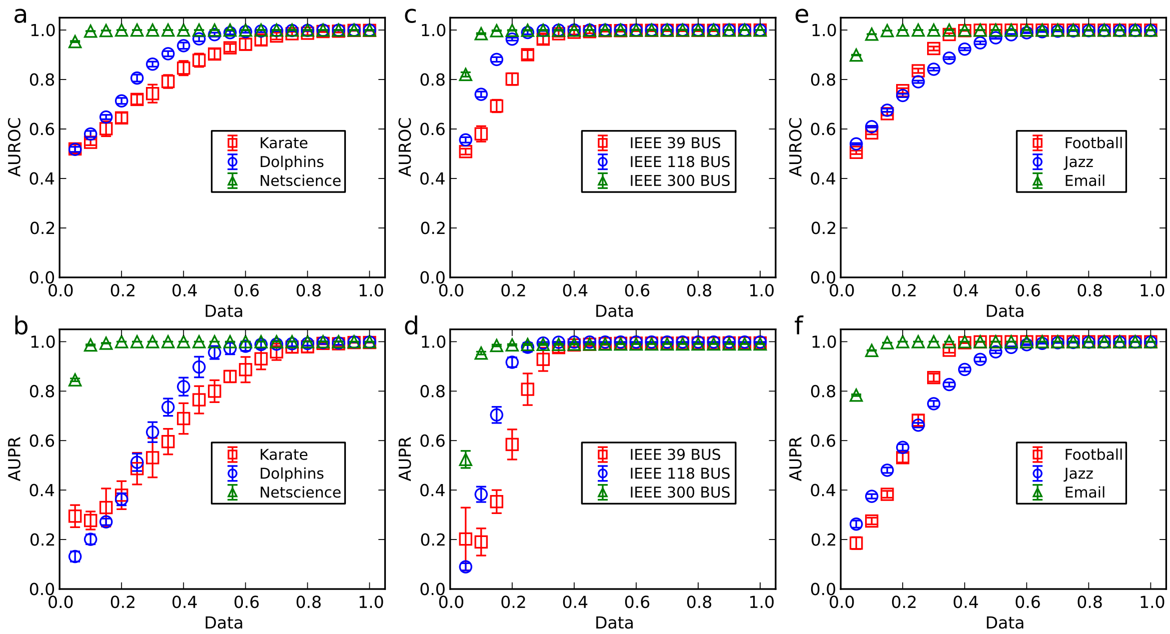
Supplementary Figure S8: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Network size N is 100 and average degree $\langle k \rangle = 18$. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Rewiring probability of SW networks is 0.3.



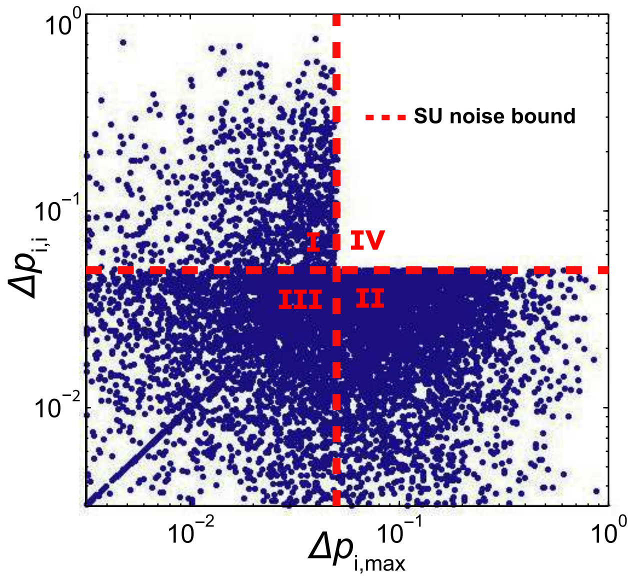
Supplementary Figure S9: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Network size N is 500 and average degree $\langle k \rangle = 6$. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Rewiring probability of SW networks is 0.3.



Supplementary Figure S10: AUROC and AUPR of reconstructing random (ER), small-world (SW) and scale-free (SF) networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Network size N is 1000 and average degree $\langle k \rangle = 6$. Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations. Rewiring probability of SW networks is 0.3.



Supplementary Figure S11: AUROC and AUPR of reconstructing several empirical networks based on the time series obtained from (a)-(b) evolutionary ultimatum games, (c)-(d) transportation of electric current and (e)-(f) communications. Three empirical networks: Karate, Dolphins and Netscience networks are used in (a)-(b). Three empirical networks, IEEE 39 BUS, IEEE 118 BUS and IEEE 300 BUS networks are used in (c)-(d). Three empirical networks, Football, Jazz and Email networks are used in (e)-(f). Each data point is obtained by averaging over 10 independent realizations. Error bars denote the standard deviations.



Supplementary Figure S12: Four phases are identified in the phase diagram: (I) imitate the best strategy in neighbors, (II) strategy unchanged, (III) indistinguishable and (IV) imitate other strategies. $\Delta p_{i,\max} \equiv |p_i(t+1) - p_{\max}(t)|$ and $\Delta p_{i,i} \equiv |p_i(t+1) - p_i(t)|$, where $|\cdot|$ represents absolute value and $p_{\max}$ is the proposal associated with maximum payoff in the neighborhood of node i. SW networks are used with network size $N = 100$, average degree $\langle k \rangle = 6$, and rewiring probability 0.3. The dashed lines are $\Delta p_{i,\max} = \delta$ and $\Delta p_{i,i} = \delta$.