TABLE OF CONTENTS

Supporting Text

S1. Classification of cyclic motifs into all their possible non-isomorphs

Complete classification of network motifs as they were previously defined for mixed graphs (graphs with both directed and undirected edges) [R Milo, S Shen-Orr, S Itzkovitz, N Kashtan, D Chklovskii, U Alon, *Network Motifs: Simple Building Blocks of Complex Networks* Science, 298:824-827 (2002)] results in the following sequence of distinct configurations for *n* nodes, where the first item in the sequence corresponds to *n*=3:

13, 199, 9364, 1530843, 880471142, 1792473955306, …

Hence, identification and classification of network motifs of sizes larger than 7 nodes is computationally impractical. We studied cycles in mixed graphs. We define cycles as closed walks that end at the node where the walk started visiting each node in the cycle exactly once regardless of the directionality of the edges. The number of distinct classes for cycles of *n* nodes (modulo reflection and rotational symmetries) with directed and undirected edges, starting with *n*=3, follows the sequence:

7, 15, 30, 74, 171, 444, 1138, 3048, 8175, 22427, 61686, 171630, …

Figure S4 shows the seven possible 3-cycle configurations and the fifteen possible 4-cycle configurations. It can be shown (see Supplement Section 1) that this sequence follows the equation:

$$N(n) = \frac{1}{2n}\sum_{i=1}^{n} 3^{\gcd(i,n)} + \begin{cases} 3^{n/2}/3 & \text{if } n \text{ is even} \\ 3^{(n-1)/2}/2 & \text{if } n \text{ is odd} \end{cases}$$

where gcd(*i*,*n*) is the greatest common divisor of *i* and *n*. We deposited this sequence into the The On-Line Encyclopedia of Integer Sequences (sequence ID A117747).

S2. Cycle search algorithm

The algorithm and pseudo-code we used to identify cycles in directed and mixed networks is the following:

Input: A graph G = (V, E)
Output: All size-k cycles in G.
for each vertexv in V do
call Expand(vertexv, currentv, k, depth, subgraph)
    endfor
Expand(startv, currentv, k, depth, subgraph)
if (current = startv and depth = k) then
output subgraph
return
elseif (current = startv and depth < k) then

```
return
    else
    for each neighborv of currentv do
            A:   if (not in subgraph and neighborv > startv) do
        add neighborv to subgraphs
        k=k+1
            Expand(startv, neighborv, k, depth, subgraph)
    endif
endfor
endif
return
```

Finding all cycles in large networks using this algorithm becomes exponentially time consuming. To deal with this problem we implemented a sampling approach where we placed a probability function before the recursive call. Line A in the above pseudocode was replaced with:

```
A:        if (not in subgraph and neighborv > startv and probability p)
```

The probability function, for the purpose of sampling, returns a probability to enter the recursive call based on a calculation that considers the network size in links and nodes, the size of the loops we are searching for, the number of processors available, using a randomly generated number. This is done in order to achieve reasonable sampling density across the entire network.

In order to take advantage of the high-performance computing capabilities of BlueGene we parallelized the program. We start the recursive expansion of the graph search from each node. If there are more nodes in the graph than computing nodes available (in the computing cluster) the code automatically divides the job optimally to use all computational resources available using the following strategy:

```
Input: Vertex v, number of processors available, processor ID, graph G
Output: YES/NO to expand vertex v
if (v MOD number of processors) == (processor ID + 1)
    return YES
else
    return NO
endif
```

We implemented three speeding enhancements to improve the efficiency of the original algorithm:

- Sorting the list of nodes by the number of links per node while placing highly connected nodes at the bottom/end of the list.
- Creating an index for fast retrieval of nodes and interactions from a linked list of nodes and a linked list of links.

Assigning a unique integer for each node and not expanding a node if its unique identifier is greater than the node we attempt to expand. This concept was introduced before by Wernicke [Sebastian Wernicke, A Faster Algorithm for Detecting Network Motifs, Lecture Notes in Computer Science, Springer Berlin / Heidelberg, ISSN 0302-9743 Pages 165-177 Volume 3692/2005, Algorithms in Bioinformatics, DOI 10.1007/11557067].

S3. Construction of network datasets

We analyzed the 9 directed networks listed below:

1) *S. cerevisiae* transcriptional regulatory network. Nodes in this network are genes/proteins affecting the transcription of other protein by binding to their DNA promoter sequence. Links indicate the binding of the former to the latter. The network was downloaded from: http://www.mrc-lmb.cam.ac.uk/genomes/madanm/tfcomb/. The network is a part of a study by Balaji et al. [Balaji S, Babu MM, Iyer LM, Luscombe NM, Aravind L. Comprehensive analysis of combinatorial regulation using the transcriptional regulatory network of yeast. J Mol Biol. 2006 Jun 30;360(1):213-27]

2) *E. coli* gene regulatory networks. The network was developed from the experimental biomedical literature and was downloaded and reconstructed from: http://regulondb.ccg.unam.mx/ (release 5.8) [Gama-Castro S, et al. RegulonDB (version 6.0): gene regulation model of Escherichia coli K-12 beyond transcription, active (experimental) annotated promoters and Textpresso navigation. Nucleic Acids Res. 2008 Jan;36(Database issue):D120-4.]

3) Mammalian cell signaling network in CA1 neurons. Nodes in this network represent proteins or small molecules and links represent direct biomolecular mass-action or enzymatic reactions. This network was developed manually from experimental biomedical literature and can be downloaded from: http://www.mssm.edu/labs/maayan/datasets/qualitative_networks.shtml. Neutral links or undirected interactions in this network represent known physical protein-protein interactions where the functional hierarchy effect (source affecting a target) between the two components is not clear or known. This is mainly reserved to scaffolding proteins and other proteins present in a complex but do not directly participate in information transfer. This network was published by our group in 2005 [Ma'ayan A, Jenkins SL, Neves S, Hasseldine A, Grace E, Dubin-Thaler B, Eungdamrong NJ, Weng G, Ram PT, Rice JJ, Kershenbaum A, Stolovitzky GA, Blitzer RD, Iyengar R. Formation of regulatory patterns during signal propagation in a Mammalian cellular network. Science. 2005 12;309(5737):1078-83].

4) *C. elegans* neuronal connectivity. Nodes in the network represent non-pharyngeal neurons and links represent three types of synaptic connections: chemical synapses, electrical junctions, and neuromuscular junctions. The network was downloaded from: http://www.wormatlas.org/handbook/nshandbook.htm/nswiring.htm. It was assembled from two original publications from 1986 and 1991 and was later updated. A full description of how thos network was created can be found at the Worm Atlas.

5) Food web of Little Rock Lake, Wisconsin. Nodes in this network are autotrophs, herbivores, carnivors or decomposers whereas links represent food sources. Downloaded from: http://www.cosin.org/extra/data/foodwebs/littlerock.txt This network was first published by Martinez in 1991 [Martinez, N. D. 1991 Artifacts or attributes? Effects of resolution on the Little Rock Lake food web. *Ecol.Monogr.* **61**, 367-392]

6) Functional brain connectivity map was created from time-series fMRI images. Nodes are voxels, and a link is established if two voxels are temporally correlated. Directionality was determined using coarse-grain causality, as described in *"Identifying directed links in large*

scale functional networks: application to brain fMRI", G.A. Cecchi, A.R. Rao, M.V. Centeno, M. Baliki, A.V. Apkarian & D.R. Chialvo, BMC Cell Biology 8(Suppl 1):S5 (2007).

7) Air traffic control network. This network was constructed from the FAA (Federal Aviation Administration) National Flight Data Center (NFDC), Preferred Routes Database. Nodes in this network represent airports or service centers and links are created from strings of preferred routes recommended by NFDC; downloaded from: http://www.fly.faa.gov

8) Electrical circuits were extracted from flat format files and benchmark electrical circuits at the gate level downloaded from: http://www.pld.ttu.ee/~maksim/benchmarks/iscas89/bench/s9234.bench

9) Tracert network. An internet connectivity map was created using the tracert command line utility from RedHat Cygwin (UNIX/Linux emulation on Windows). Requests were sent from a host in New Jersey, US to randomly generated web-sites containing the following template: www.[A-Z][A-Z][A-Z].com.

All nine networks are provided as SOM in the file **network.xls** where interactions are listed in 3 columns: the first column is for source nodes, the second column for target nodes, and the third column for the effect. If the link has direction the effect is either positive (+) or negative (_) it is included where the default is positive (+). We did not consider the sign of the effect for most of the analyses in this study. If the link does not have directionality it is considered neutral (0). With neutral interactions the designation of source and target nodes is not considered such that the source nodes can be the targets and the targets can be the sources.

S4. Spin Model for the computation of the magnetization in cyclic motifs

We define the Hamiltonian for a 3-state spin system with nearest-neighbor interactions as $\beta H = -J\sum_{i=1}^{N} s_i s_{i+1} - h\sum_{i=1}^{N} s_i$ where a circular topology is assumed, $s_{i+N} = s_i$, h is an external field, J is the coupling interaction between spins, $s_i = \{-1,0,1\}$ and $\beta = 1/kT$. The parameters J and h are measured in units of kT. Notice that we assume that neutral links (corresponding to s_i=0) do not interact with directed links or among themselves. To control for the varying density of undirected links we include a chemical potential term μ (also measured in units of kT), so that the Hamiltonian is extended to:

$$H = -J\sum_{i=1}^{N} s_i s_{i+1} - h\sum_{i=1}^{N} s_i - \mu\sum_{i=1}^{N}(1 - s_i^2) \qquad \text{(S1)}$$

The partition function $Z = \sum_{\{s_i\}} e^{-\beta H(\{s_i\})}$ can be computed analytically using the transfer-matrix method:

$$K = \begin{pmatrix} e^{h+J} & e^{h/2+\mu/2} & e^{-J} \\ e^{h/2+\mu/2} & e^{\mu} & e^{-h/2+\mu/2} \\ e^{-J} & e^{-h/2+\mu/2} & e^{-h+J} \end{pmatrix} \qquad \text{(S2)}$$

$$Z = tr\{K^N\} = \lambda_1^N + \lambda_2^N + \lambda_3^N \qquad \text{(S3)}$$

where λ_i are the eigenvalues of K. From the partition function we can derive the first and second moments of the magnetization in the absence of external magnetic field:

$$\langle M \rangle = \left\langle \sum_{i=1}^{N} s_i \right\rangle = \frac{\partial \ln Z}{\partial h} \qquad \text{(S4)}$$

$$\langle M \rangle (h=0) = 0 \qquad \text{(S5)}$$

$$\langle M^2 \rangle (J, \mu, N, h=0) == \frac{\partial^2 \ln Z}{\partial h^2} \qquad \text{(S6)}$$

as well as the average number of neutral links:

$$\langle L \rangle = \left\langle \sum_{i=1}^{N} (1 - s_i^2) \right\rangle; \qquad \text{(S7)}$$

$$\langle L \rangle (J, \mu, N, h=0) = \frac{\partial \ln Z}{\partial \mu}. \qquad \text{(S8)}$$

Henceforth, we will omit $h=0$ and simply refer to the second moment of the magnetization and mean number of $s=0$ spins with $M^2 = M^2(J, \mu, N)$ and $L = L(J, \mu, N)$. In order to test the validity of this approach, we estimated a unique pair of parameters (J_k, μ_k) for each network k, by minimizing the error function:

$$\varepsilon_k = \sum_N \left[\langle M_k^2(N) \rangle - M^2(J_k, \mu_k, N) \right]^2 + \sum_N \left[\langle L_k(N) \rangle - L(J_k, \mu_k, N) \right]^2 \qquad \text{(S9)}$$

w.r.t J_k and μ_k. Notice that N spans a range between 3 and ~15, and therefore this is a two parameter fit to ~12 data points The measurements for the biological networks are as follows: the network equivalent of the second moment of the magnetization for each cycle of length N is computed as $M^2 = \left\langle (n_l - n_r)^2 \right\rangle$, where $n_{l,r}$ is the total number of clockwise and counterclockwise arrows along the cycle; the density equivalent L_k is simply the ratio of undirected links over total links for the entire network and is independent of N.

As the eigenvalues of the transfer matrix can be derived analytically and computed using standard mathematical packages, it is possible to evaluate directly $\langle M^2 \rangle$ and $\langle L \rangle$ as a function of J, μ and N; these analytic expressions were used to find the values of J_k and μ_k (see Table 1) that minimize the error ε_k. In order to simplify the information presented in Fig. 2, we used the values of J and μ obtained as explained above for each network, and computed and displayed the easier to interpret quantity $\langle | M | \rangle$ sampling through a classical Monte Carlo method, instead of displaying $\langle M^2 \rangle$.

S5. Derivation of the number of non-isomorphic N-cycles

We consider cyclic graphs, with three types of links: neutral (i.e., no arrow), clockwise arrow and counterclockwise arrow (see Fig. 1 in the main text and Fig. S1).

Notation: We denote neutral links with a "0", clockwise links with a "1" and counterclockwise links with a "-1", as indicated in Fig. S1.

Convention: Given a cycle, we write its sequence of links in a clockwise fashion. For example, the cycle in Fig. S1 can be written as '0 -1 1 1' if we start from node A.

Symmetries The following symmetry operations take a graph to the same graph, even if its sequence representation is different.

Rotations: Given a sequence s, a rotation $R_k(s)$ changes the origin of s to the k-th node from the origin going clockwise. s = 0 -1 1 1 → $R_1(s)$ = -1 1 1 0 → $R_2(s)$ = 1 1 0 -1 → $R_3(s)$ = 1 0 -1 1

Flips: Given a sequence s, a flip F(s) changes the order in which we read the cycles from clockwise to counterclockwise, simultaneously changing the direction of the directed arrows (see Fig. S2): s = 0 -1 1 1 → F(s) = -1 -1 1 0.

Note that arrows that are clockwise in the original graph become counter-clockwise in the flipped graphs. Therefore a flip operation is written by flipping the original sequence '0 -1 1 1' from right to left, and changing -1 into 1, 1 into -1 and leaving 0 unchanged. Clearly F(F(s))=s.

Special cycles

Periodic cycles A periodic cycle s with *n* nodes is a cycle such that there exists a $p < n$ such that $R_p(s) = s$. For example: s = '1 -1 0 -1 1 -1 0 -1' is periodic with period 4, because $R_4(s) = s$. The largest possible period contained in a sequence $s_1\ s_2\ s_3\ \ldots\ s_n$ such that one period starts at position 1 and the same period or a subsequent one ends at position *k* is the gcd(k,n), where gcd(k,n) denotes the greatest common divisor between *k* and *n*. There are 3^p possible sequences with period *p* for cycles whose length *n* is a multiple of *p*. These include all periodic sequences with period *q* such that *q* is a divisor of *p*.

Flip invariant cycles A flip invariant cycle is a cycle that is invariant under flips, that is s = F(s). For example: s = '1 1 0 -1 -1' is flip invariant because it verifies that F(s) = s.

Rotation-flip invariant cycles. We call a cycle *s* rotation-flip invariant, if there exists a rotation of s that leads to a cycle that is the flipped version of the original one. In other words, s is rotation-flip invariant if there exists a *k* such that $R_k(s) = F(s)$. For example: s = '1 -1 -1 1 0 -1 1' ; $R_1(s)$ = '-1 -1 1 0 -1 1 1'; $R_2(s)$ = '-1 1 0 -1 1 1 -1' = F(s).

Count of the number of rotation-flip invariant cycles.

The number of possible rotation flip invariant cycles depends on the parity of the length *n* of the cycle. If *n* is odd the condition $R_k(s) = F(s)$ imposes some restrictions on the sequence. Take n = 7 and k =3. Denoting $\underline{s_i} = -s_j$, the rotation-flip invariant constraints for sequence s = s1 s2 s3 s4 s5 s6 s7 for k=3 are shown in Table S1. From Table S1 we see that the constraint in this case is that s2 = 0, given that s2 = -s2. The same would have happened if *k* = 0 (in which case s4 = 0), or *k* = 1 (in which case s3 = 0), etc. Therefore, for odd *n* the number of flip invariant cycles with a rotation at which the invariance obtains has only (n-1)/2 degrees of freedom. In effect, one of the elements has to be 0, while one half of the remaining n-1 elements are determined by the other (n-1)/2 values. As each of the degrees of freedom can take 3 values, the number of Rotation-flip invariant *n*-cycles for odd *n* results to be: $3^{(n-1)/2}$

On the other hand, if *n* is even the condition $R_k(s) = F(s)$ imposes other restrictions on the sequence, depending on the parity of *k*. Take *n* = 6 and *k* =3 and *k*=4. Again, if denoting $\underline{s_i} = -s_j$, the constraints for sequence s = s1 s2 s3 s4 s5 s6, and k=3 are shown in Table S2. From Table S2 we can see that the constraints are s2 = 0 and s5 =0. However, for *k*=4, no constrain is imposed. In general, for any even k=0,2,…,n, no constraint is imposed, whereas for odd k = 1, 3, …, n-1, two values of the sequence have to be zero. Therefore, for even *n* the number of flip invariant cycles with the rotation *k* at which

the invariance obtains has only (n-2)/2 degrees of freedom if k is odd and n/2 degrees of freedom if k is even..

Let's concentrate now on the even k case. For every even k, there are $3^{n/2}$ sequences. For each of these sequences for a given k, there is an equivalent one (via a translation) in the set corresponding to another k. Furthermore, except for the all-zero sequence, every sequence has a translation invariant within the same set. For example, take n = 4 and k = 2, the corresponding constraints are shown in Table S3, and the resulting group of sequences is shown in Table S4. So it can be seen (and proved in general) that for the chosen representative of even k, the number of unique (not reachable to each other through rotations) is $(3^{n/2}-1)/2$, where the subtraction of 1 stems from the sequence Z of all zeros.

In a similar way, it can be seen (and proved in general) that for the chosen representative of odd k, the number of unique sequences (i.e., not reachable to each other through rotations) is $(3^{(n-2)/2}-1)/2$, where again the subtraction of 1 stems from the sequence of all zeros.

Finally, the number of unique rotation-flip invariant cycles is the sum of the ones reached through the odd and even k, plus the sequences of zeros, that is: Number of Rotation-flip invariant cycles = $3^{(n-2)/2}/2+3^{n/2}/2=(2/3) \times 3^{n/2}$

Count of the total number of non-equivalent directed cycles.

Let's call N_n the number of non-equivalent cycles of length n. We first divide N_n into two sets: a set S_1 that is composed of all the sequences that are non flip-rotation invariant and a second set S_2 that contains only the sequences that are flip-rotation invariant, as shown in the Fig. S3.a. Next we create all the Rotations of these sequences, and all the flips of these sequences, leading to the four rows of boxes in Fig. S3b.

We will argue that the number of sequences in the first three rows of boxes is equal to $\sum_{j=1}^{n} 3^{\gcd(j,n)}$. To see this, let's work out a concrete example with n = 4. We will organize our count in order of periodic cycles, starting with cycles of period 1 whose repeat subsequence ends at j=1. There are three such sequences ($3^{\gcd(1,4)}$) listed in Table S5 (the first 3 rows of the first column, noted by [j,p], where j=1 is the working offset in the sequence, and p=1 indicate the period).

Let's continue with the periodic sequences that have a period that start at location 1 and a longest period (which may be different form the one starting at location 1) ending at j=2. These sequences are period 2 sequences (which include all period 1 sequences), are shown in Table S5, and noted with [j,p] = [2,2]. Note that some of the resulting 9 (i.e., $3^{\gcd(2,4)}$) sequences, some are simply Rotations of the sequences of period 1, or Rotations of other the sequences of period 2.

Next we consider j=3 and list the sequences that have the longest period that end at j=3 and start at j=1. As n=4 cannot accommodate sequences with period 3, and the sequences of period 2 cannot have a period that start at 1 and end at j=3, the only possibility are the periodic sequences of period 1, of which there are 3 ($3^{\gcd(3,4)}$). Those are annotated with a [3,2] next to them. We list these sequences in the first three rows, considering them rotations of the period 1 sequences discovered for lower values of j.

Finally, we have the sequences whose longest period end at location 4, and start at location j=1. These are all periodic sequences of period 1, period 2 and period 4. Period 4 is not really a period, because it encompasses all the length of the sequence, but we will take here as a convention that if a period appears exactly k times in the whole sequence ($k \geq 1$) we will consider that sequence periodic. (E.g., for n=4 and j=4, k=1.) There are $3^{\gcd(4,4)}$=81 sequences with period 4, out of which we have already

seen the 9 sequences with period 2 (which include the 3 sequences with period 1). Those will be located in the same row where we located them when we first found them with a lower period, but in subsequence columns. This completes the first 6 rows of Table S5. We have to still add the remaining 72 sequences with period 4 that we haven't added yet. (We will not do it explicitly here.) Once we have added all these sequences, we would have completed the three rows of boxes of Fig. 3b. But to complete the count of all the sequences in Fig. 3b, including the fourth row, we still need to add one more time the flip-rotation invariant sequences that appear just once in the consideration of the periodic sequences.

We can now assemble the puzzle. The sum $\sum_{j=1}^{n} 3^{\gcd(j,n)}$ accounts for the upper rows of boxes in Fig. 3b. But we still have to add the count of flip-rotation invariant sequences (computed earlier on). This count will have to be multiplied by n, given that we have to include all the rotations for those sequences. In this way, we obtain:

$$2nN_n = \sum_{j=1}^{n} 3^{\gcd(j,n)} + n\begin{cases} 3^{(n-1)/2} & \text{if } n \text{ is odd} \\ \frac{2}{3}3^{n/2} & \text{if } n \text{ is even} \end{cases} \qquad \text{(S10)}$$

which leads to our final result:

$$N_n = \frac{1}{2n}\sum_{j=1}^{n} 3^{\gcd(j,n)} + \begin{cases} \frac{1}{2}3^{(n-1)/2} & \text{if } n \text{ is odd} \\ \frac{1}{3}3^{n/2} & \text{if } n \text{ is even} \end{cases} \qquad \text{(S11)}$$

Table S6 shows the first 15 elements of the N_n sequence. Figure S4 shows the 7 different configurations of 3-node-cycles and the 15 configurations of 4-node-cycles.

S6. Derivation of the magnetization for the case of a network with independent edges

For an *n*-cycle with random assignment of directed edges and no neutral edges, the magnetization can be written in terms of the spins assigned to the edges as:

$$M_n = \sum_{j=1}^{n} s_j = m_+ - m_- , \qquad \text{(S12)}$$

where m_+ and m_- are the numer of the counterclockwise (spin 1) and clockwise (spin -1) edges in the cycle, respectively. As $m_+ + m_- = n$, we have that

$$|M_n| = \begin{cases} n - 2m_+ & if\ m_+ \le n/2 \\ 2m_+ - n & if\ m_+ > n/2 \end{cases} \qquad \text{(S13)}$$

where *n* is the length of the cycle. The probability of having m_+ counterclockwise edges in a cycle of length n is given by

$$P(m_+) = \binom{n}{m_+} \frac{1}{2^n}. \qquad \text{(S14)}$$

Case *n* even and no neutral links in the *n*-cycle. We will denote the conditional average of the absolute value of the magnetization given that we have no neutral links by $\langle |M_n| \,|\, 0 \rangle$. Writing $n=2k$, we have that

$$\begin{aligned} \langle |M_n| \,|\, 0 \rangle &= \sum_{m_+=0}^{2k} |\, 2k - 2m_+ \,|\, P(m_+) \\ &= 2\sum_{j=0}^{k} (k-j) P(j) + 2\sum_{j=k+1}^{2k} (j-k) P(j) \\ &= 4\sum_{j=0}^{k} (k-j) \binom{2k}{j} \frac{1}{2^{2k}} \end{aligned} \qquad \text{(S15)}$$

That is

$$\langle |M_n| \,|\, 0 \rangle = \frac{4}{2^{2k-1}} \left[\sum_{j=0}^{k} k \binom{2k}{j} - \sum_{j=0}^{k} j \binom{2k}{j} \right] \qquad \text{(S16)}$$

To find an easier expression for Equation S16, we re-work the two sums between brackets in Eq. (S16). For the first sum, we observe that

$$\begin{aligned} 2^{2k} &= \sum_{j=0}^{2k} \binom{2k}{j} \\ &= \sum_{j=0}^{k} \binom{2k}{j} + \sum_{j=k+1}^{2k} \binom{2k}{j} \\ &= \sum_{j=0}^{k} \binom{2k}{j} + \sum_{j=0}^{k-1} \binom{2k}{2k-j} \\ &= 2\sum_{j=0}^{k} \binom{2k}{j} - \binom{2k}{k} \end{aligned} \qquad \text{(S17)}$$

From where the first sum between brackets in Eq. (S16) results:

$$\sum_{j=0}^{k} k \binom{2k}{j} = k \left[2^{2k-1} + \frac{1}{2} \binom{2k}{k} \right]. \qquad \text{(S18)}$$

For the second sum between brackets in (S16), we use the identity $j\binom{2k}{j} = 2k\binom{2k-1}{j-1}$, and compute

$$\sum_{j=0}^{k} j\binom{2k}{j} = 2k\sum_{j=1}^{k}\binom{2k-1}{j-1} = 2k\sum_{j=0}^{k-1}\binom{2k-1}{j} \tag{S19}$$

Making use of the identity

$$2^{2k-1} = 2\sum_{j=0}^{k-1}\binom{2k-1}{j}, \tag{S20}$$

we find that

$$\sum_{j=0}^{k} j\binom{2k}{j} = k2^{2k-1} \tag{S21}$$

Replacing (S18) and (S20) in (S16), we obtain for even *n* that

$$\left\langle |M_n| \right| 0\right\rangle = \frac{2k}{2^{2k}}\binom{2k}{k}; \qquad k=n/2 \tag{S22}$$

Case *n* odd and no neutral links in the *n*-cycle. Writing $n=2k-1$, we have that the equivalent of (1) is

$$\left\langle |M_n| \right| 0\right\rangle = \frac{2}{2^{2k-1}}\left[\sum_{j=0}^{k-1}(2k-1)\binom{2k-1}{j} - \sum_{j=0}^{k-1} 2j\binom{2k-1}{j}\right] \tag{S23}$$

Using the identity $j\binom{2k}{j} = 2k\binom{2k-1}{j-1}$, we can rearrange the second sum in the square brackets of (S23) as

$$\sum_{j=0}^{k-1} j\binom{2k-1}{j} = (2k-1)\sum_{j=1}^{k-1}\binom{2k-2}{j-1} = (2k-1)\sum_{j=0}^{k-2}\binom{2k-2}{j} \tag{S24}$$

and making use of Eq. (S18) we obtain that

$$\sum_{j=0}^{k-1} j\binom{2k-1}{j} = (2k-1)\frac{1}{2}\left(2^{2k-2} - \binom{2k-2}{k-1}\right) \tag{S25}$$

Replacing (S25) and (S20) in (S23), we obtain

$$\left\langle \left| M_n \right| \mid 0 \right\rangle = \frac{2k}{2^{2k}} \binom{2k}{k}; \qquad k=(n+1)/2 \tag{S26}$$

Eqns. (S22), valid for even *n*, and (S26), valid for odd *n*, can be summarized in the general formula, valid for all $n>0$:

$$\left\langle \left| M_n \right| \mid 0 \right\rangle = \frac{2k}{2^{2k}} \binom{2k}{k}; \qquad k = \mathrm{int}\left(\frac{n+1}{2} \right). \tag{S27}$$

Case any n and finite number of neutral links. Equation (S27) is valid when there are no neutral edges in the *n*-cycles cycle. If there are ν neutral edges in a cycle, the previous equation remains valid, provided that we replace ν by n- ν. If we denote the conditional average of the absolute value of the magnetization of an *n*-cycle given that it has exactly ν neutral edges by $\left\langle \left| M_n \right| \mid \nu \right\rangle$, then it is clear that

$$\left\langle \left| M_n \right| \mid \nu \right\rangle = \left\langle \left| M_{n-\nu} \right| \mid 0 \right\rangle. \tag{S28}$$

If the fraction of neutral links in the whole network *p*, then the number of neutral links in each cycle will be variable, presumable following a binomial distribution with parameters *n* and *p*. Therefore, the average of absolute value of the magnetization under these conditions is

$$\begin{aligned} \left\langle \left| M_n \right| \right\rangle &= \sum_{\nu=0}^{n} \left\langle \left| M_n \right| \mid \nu \right\rangle \binom{n}{\nu} p^{\nu} (1-p)^{n-\nu} \\ &= \sum_{j=0}^{n} \left\langle \left| M_j \right| \mid 0 \right\rangle \binom{n}{j} p^{n-j} (1-p)^{j} \end{aligned} \tag{S29}$$

Finally, we can find an asymptotic expression for Eq. S27 for $n>>1$. Using Stirling formula it can be shown that the central binomial coefficient is

$$\binom{2k}{k} = \sqrt{\frac{2}{\pi}} \frac{2^{2k}}{\sqrt{2k+1}} \tag{S30}$$

and therefore

$$\left\langle \left| M_n \right| \mid 0 \right\rangle \cong \sqrt{\frac{2}{\pi}} \frac{2k}{\sqrt{2k+1}} = \sqrt{\frac{2}{\pi}} \begin{cases} \dfrac{n+1}{\sqrt{n+2}} \approx \sqrt{n} & \text{if odd } n >>1 \\ \dfrac{n}{\sqrt{n+1}} \approx \sqrt{n} - \dfrac{1}{2\sqrt{n}} & \text{if even } n>>1 \end{cases} \tag{S31}$$

S7. Statistics of the edge-edge correlation ρ

To quantify the number of different types of nodes in a directed cycle, we introduced the edge-edge correlation ρ for a node, which by definition takes the values "-1" for sinks and sources, "1" for pass-through nodes and "0" for nodes with at least one neutral/ bidirectional link. In the spin system representation of cycles, each spin represents an edge with values 0, -1 or 1 depending on whether the edge is neutral or traverses the cycle in a clockwise or counterclockwise direction. Therefore

$$\rho = s_i s_{i+1}.$$

Figure S5 shows the distribution of the edge-edge correlation for the different networks and for all the cycle lengths considered. Except for the FAA, all the networks exhibit a tendency to have more weight in the ρ=-1 bin, consistent with the existence of anti-ferromagnetic interactions. The abundance of sink and source nodes in cycles changed with cycle length for the neuronal network, the internet, and the electrical-circuit. For other networks, such as the ecosystem foodweb, brain and FAA, ρ remained constant for all loop sizes.

S8. The exclusion principle

One possible explanation for the observed anti-ferromagnetism could be the existence of hubs that contain almost exclusively either incoming or outgoing edges. We call this possible organization of in and out degree distributions the "exclusion principle". Clearly, if an exclusion principle is rigorously at play, a cycle passing through a node with an incoming arrow, should "exit" the node with an incoming arrow as well, making the node a sink for that cycle. (A similar description would apply for sources.) A measure of exclusion per node x can be defined as the difference between the in-degree d_{in} and out-degree d_{out} of a node, normalized by the maximum between d_{in} and d_{out} (neutral links are not included in this calculation). For nodes with a perfect exclusion principle, we would have either x=1 (pure sink nodes) or x= -1 (pure source node), and the average $\langle x^2 \rangle$ (which we call exclusion degree) over all the nodes in the network would be 1. Alternatively, if the exclusion principle didn't hold and the x-distribution was uniform across all the nodes, then we expect $\langle x^2 \rangle = 0.33$. Unimodal distributions with the mode of x between -1 and 1 will have $\langle x^2 \rangle < 0.33$.

Figure S6 shows the histograms of x for the hub nodes in the upper 10 percentile of the directed degree distribution. If an exclusion principle were at play, we would observe bimodal distributions. This is indeed the case for some, but not all the networks. For example, tracert has an exclusion degree of 0.27. Still, the tracert network has a parameter J ~-1 (see Table 1), indicating a clear anti-ferromagnetic interaction between edges. In Figure S7 we take another look at the exclusion principle, which includes all the nodes in the networks (and not just the 10% most-connected nodes). We can see a tendency for the nodes with largest in-degree or out-degree to align in the diagonal of the plots, indicating that hubs have a high exclusion degree. The thickening of the cloud of points towards smaller in- or out-degree shows that the exclusion principle is less valid for nodes with intermediate to low values of in and out degrees.

S9. Relationship between the exclusion principle and the magnetization

The scatter graph of the coupling constant J versus the exclusion degree $\langle x^2 \rangle$ (Fig. S8a) shows that there is a substantial correlation (correlation coefficient r = -0.74) between J and $\langle x^2 \rangle$.

The correlation is a little stronger (r = -0.82) in the region $<x^2>>0.75$ (Fig. S8b). Hence, if $<x^2>$ >0.75 then the coupling constant J < -0.5. As a rule of thumb we will say that the exclusion principle holds if $<x^2>$ > 0.75. Following this, we conjecture that the Exclusion Principle leads to Anti-ferromagnetism.

However, the horizontal rectangle in Fig. S8a contains points such that J < -0.5 (which is anti-ferromagnetic), but the exclusion degree is violated for some of those points. Therefore, we conclude that the exclusion principle is a sufficient (but not necessary) condition for anti-ferromagnetism.

Example where the Exclusion Principle leads to Anti-ferromagnetism
If the exclusion principle were perfectly observed ($<x2>$=1), then the magnetization of each cycle would be minimal. The reason is that if all nodes are either sinks or sources in the original graph (Fig. S8c), then they will be sink or sources in any cycle that they participate in. Notice also that in this case, there would be no cycle with an odd number of edges.

Examples where Anti-ferromagnetism does not leads to Exclusion Principle
If the exclusion principle were violated, ($<x2>$ < 0.75), we show that the magnetization of each cycle could still be minimal. The reason is that even if there are few sinks and sources, the global architecture of the graph can allow for low magnetization cycles. Figure S8d an example of one cycle of length 4 (with magnetization = 0) and four cycles of length 3 (with magnetization = 1, i.e., the minimal for off length cycles). The exclusion degree in this case is 0.33, but we still have anti-ferromagnetism (i.e, the magnetization is the minimal it can be).

S10. The magnetization of real vs. randomized networks

In Figure 2(e) and (f) of the main paper we plotted the magnetization for the foodweb and signaling networks, the in-and-out-degree-preserving (IODP) randomization (red lines), the total-degree-preserving (TDP) randomization (green line). If Fig. S9 we show the equivalent plots for all the 9 networks. In all the cases the actual network magnetization is smaller than their randomized counterparts, except for the FAA network, for which there is no difference between actual and randomized networks. The magnetization pertaining to the TDP randomization is perfectly explained by the analytical formula derived earlier in this Supplement.

S11. Linear Stability Analysis of cycles and feedback loops

To analyze the linear stability of the dynamics on cyclic topologies, we will consider cycles with directed links, excluding the case of non-directed/bidirectional links. In such cases, the number of sinks is equal to the number of sources. We will separate the case of cycles with at least one source and the case of cycles with no sources (and therefore no sinks). The latter case corresponds to feedback loops.

Linear stability analysis of non-feedback loop cycles. Consider an n-cycle, with k>1 sources. We will assign indexes to the nodes as follows. The first k indexes will enumerate the sources, in no particular order. Identifying the pass-through nodes as clockwise (j nodes) and counterclockwise (l nodes), we index the j+l pass-through nodes as follows. Starting from source node number 1, and traversing the cycle in the clockwise direction, we index the clockwise pass-through nodes with indexes k+1, k+2,…, k+j. In a similar way, starting from source node number 1, and now traversing the cycle in the counter-clockwise direction, we index the counter-clockwise pass-through nodes with indexes k+j+1, k+j+2,…,

k+*j*+*l*. The remaining *k* sinks in the cycle are assigned indexes *n*-*k*+1,…,*n*. (See Fig. 3a in the main text for an example of this indexing scheme.) With this index convention, the dynamical equations for the sources are of the form

$$\frac{d}{dt}x_i = -\delta_i x_i, \quad 1 \le i \le k. \qquad \text{(S30)}$$

Note that the dynamic equations of the sources are autonomous. We now turn to the dynamic equations for the pass-through nodes. Because of the choice of indexes, each pass-through node is activated either by a pass-through node with a lower index or by a source node. That is

$$\frac{d}{dt}x_i = -\delta_i x_i + \alpha_{ir} x_r, \quad r < i, \quad k < i \le n-k. \qquad \text{(S31)}$$

Finally, the sink nodes are activated by either pass-through nodes or by sources whose indexes are smaller that the indexes of the sink nodes. The sink node equations are

$$\frac{d}{dt}x_i = -\delta_i x_i + \alpha_{ir} x_r + \alpha_{it} x_t, \quad r,t < i, \quad n-k < i \le n. \qquad \text{(S32)}$$

In all the equations above, α_{rs} represents the strength with which node *s* controls node *r*, and arise from the linearization of the (possibly non-linear) dynamical equations around a fixed point that without loss of generality can be assumed to be the origin (see next SOM section: Dynamical systems analyses of directed networks). The α's can be positive or negative. The δs represents the degradation rate and are positive constants. In this way, a linear dynamics on this cycle would be represented by an equation of the form

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix} = A \begin{bmatrix} x_1 \\ x_2 \\ \dots \\ x_n \end{bmatrix}; \qquad \text{(S33)}$$

The indexing conventions established above forces A to be a lower triangular matrix of the form

$$A = \begin{bmatrix} -\delta_r & 0 \\ \alpha_{rs} & -\delta_s \end{bmatrix} \qquad \text{(S34)}$$

The linear stability of this system of ordinary differential equations is determined by the real part of the eigenvalues of the matrix A. If the maximum real part of the eigenvalues is positive, then the system is linearly unstable, and small perturbations around the fixed point will grow exponentially. Conversely, if the largest of the real part of the eigenvalues of matrix A is negative, then the system is linearly stable, and small perturbations around the fixed point will subside, bringing the dynamical state of the system back to the fixed point. As the matrix A for cycles with sinks and sources is lower triangular, its eigenvalues are given by the diagonal elements $-\delta_1$, $-\delta_2$, …, $-\delta_n$, and therefore non-feedback loop cycles are linearly stable for all values of the parameters δ's and α's, as stated in Figure 3a.

Linear stability analysis of feedback loop cycles. Consider an *n*-cycle, with *n* pass-through nodes. We will assign indexes to the nodes as follows. Choosing an arbitrary node as node 1 we index the nodes in increasing order as we traverse the cycle in the direction of the links. (See Fig. 3b in the main text for an example of this indexing scheme.) As usual, the linearized dynamical equations are

$$\frac{d}{dt}\begin{bmatrix} x_1 \\ x_2 \\ x_i \\ x_n \end{bmatrix} = A\begin{bmatrix} x_1 \\ x_2 \\ x_i \\ x_n \end{bmatrix} \qquad \text{(S35)}$$

where, because of the index convention we chose above, the matrix A is of the form

$$A = \begin{bmatrix} -\delta & 0 & 0 & \alpha_n \\ \alpha_1 & -\delta & 0 & 0 \\ 0 & \alpha_{i-1} & -\delta & 0 \\ 0 & 0 & \alpha_{n-1} & -\delta \end{bmatrix} \qquad \text{(S36)}$$

For simplicity, we have assumed that all δs are all the same, but this assumption can be removed without much change in the conclusions. As before, the stability of this system of equations depends on the eigenvalues ξ of the matrix A, which are computed as the roots of the characteristic polynomial $p(\xi) = \det\left[A - \xi \mathbf{1}\right]$. This determinant can be expanded along the first row of the matrix, to yield

$$p(\xi) = (-1)^n \left\{ \left(\xi + \delta\right)^n - \prod_{i=1}^{n} \alpha_i \right\}. \qquad \text{(S37)}$$

There are two cases to consider. The first case is the case of positive feedback loops, in which $\prod_{i=1}^{n} \alpha_i > 0$. In this case the *n* roots of $p(\xi)$ are (taking into account the *n* roots of unity)

$$\xi_j = -\delta + \left(\prod_{i=1}^{n} \alpha_i \right)^{1/n} \left\{ \cos\left[j\frac{2\pi}{n} \right] + i \sin\left[j\frac{2\pi}{n} \right] \right\}, \qquad j = 0, ..., n-1 \qquad \text{(S38)}$$

and the maximum real part, corresponding to the most unstable eigenvalue, is

$$\lambda \equiv \max\left\{ \mathrm{Re}\left(\xi_j\right) \right\} = -\delta + \left(\prod_{i=1}^{n} \alpha_i \right)^{1/n}. \qquad \text{(S39)}$$

The second case is the case of negative feedback loops, in which $\prod_{i=1}^{n} \alpha_i < 0$. In this case the *n* roots of $p(\xi)$ are (taking into account the *n* roots of -1)

$$\xi_j = -\delta + \left|\prod_{i=1}^{n}\alpha_i\right|^{1/n}\left\{\cos\left[(2j+1)\frac{\pi}{n}\right] + i\sin\left[(2j+1)\frac{\pi}{n}\right]\right\} \qquad \text{(S40)}$$

and the maximum real part, corresponding to the most unstable eigenvalue, is

$$\lambda \equiv \max\left\{\mathrm{Re}\left(\xi_j\right)\right\} = -\delta + \left|\prod_{i=1}^{n}\alpha_i\right|^{1/n}\cos\left(\frac{\pi}{n}\right) \qquad \text{(S41)}$$

For both negative and positive feedback loop cases, the condition of stability $\lambda < 0$, which translates into

$$\frac{\left|\prod_{i=1}^{n}\alpha_i\right|^{1/n}}{\delta} < \begin{cases} 1 & \text{if } \prod_{i=1}^{n}\alpha_i > 0 \\ \dfrac{1}{\cos(\pi/n)} & \text{if } \prod_{i=1}^{n}\alpha_i < 0 \end{cases} \qquad \text{(S41)}$$

These are the stability conditions of Fig. 3b in the main text.

S12. Dynamical systems analyses of directed networks

Assuming that each node has a dynamic variable associated with it, e.g., the concentration of the corresponding protein, we envision a dynamical system of the form $\frac{dx_i}{dt} = -\delta x_i + f_i(x_1,...,x_n) + b_i + \varepsilon\xi_i(t)$, for $1 \le i \le n$, where we assume that each molecular species has a uniform degradation rate δ and a production rate b_i. The interaction between the different species is given by the vector field $f_i(x$'s) and each species is affected by a white-in-time noise ξ_i of magnitude given by ε. If the noise intensity ε were zero, the system would reach a stationary state $\mathbf{x}^0$ that balance the production, degradation and interactions, and which verifies the equations $-\delta x_i^0 + f_i(\mathbf{x}^0) + b_i = 0$. As is customary, we assume that the deviations away from $\mathbf{x}^0$ are not large and linearize the system of equations around $\mathbf{x}^0$, to obtain $\frac{dy_i}{dt} = -\delta y_i + \sum_{j=1}^{n}\left.\frac{\partial f_i}{\partial x_j}\right|_{\mathbf{x}^0} y_j + \varepsilon\xi_i(t)$, where $y_i = x_i - x_i^0$. We will also assume that the entries of the Jacobian matrix $\frac{\partial f_i}{\partial x_j}$ are either zero (i.e., f_i does not depend on x_j at $\mathbf{x}^0$) or are of the same order of magnitude given by the parameter α. Modeled in this way, the Jacobian matrix can be written as $\frac{\partial f_i}{\partial x_j} = \alpha\, s_{ij}\, A_{ij}^T$, where A^T is the transpose of the adjacency matrix A (whose the element ij is equal to 1 if a directed edge goes between node i and node j and zero otherwise), and $s_{ij} = \mathrm{sgn}(\frac{\partial f_i}{\partial x_j})$. We will call $s_{ij}\, A_{ij}$ the signed adjacency matrix, and will denote it by $\tilde{A}$. In this way, if we use $1/\delta$ as the unit of time we obtain that the linearized dynamic equations are given by $\frac{dy_i}{d\tau} = -y_i + \lambda\sum_{j=1}^{n}\tilde{A}_{ij}^T y_j + \sigma\xi_i(\tau)$, where we called $\tau = \delta t$, $\lambda = \alpha/\delta$ and $\sigma = \varepsilon/\delta$. To avoid divergence of the solutions of these linearized dynamic equations, we modify the dissipative term into $-\sinh(y_i)$

which coincides with the original $-y_i$ for $|y_i| << 1$, but creates a nonlinear saturation that avoids divergences for $|y_i| >> 1$. The resulting dynamic equations are:

$$\frac{dy_i}{d\tau} = -\sinh(y_i) + \lambda \sum_{j=1}^{n} \tilde{A}_{ij}^{T} y_j + \sigma \xi_i(\tau) \qquad \text{(S42)}$$

Therefore the previous equations capture the dynamics of directed networks in a general sense. The linear term describing the local dynamical properties of each node, $\tilde{A}$ is the signed adjacency matrix describing the connectivity, and λ is the parameter that determines the relative effects of the interactions as compared with the degradation rate. While these equations model the network dynamics around a fixed point, large deviations from the fixed points (if they exist) do not produce divergencies due to the dissipative nature of the sinh term. Therefore, these equations can capture the effects that linear instabilities have on the collective dynamics of coupled nodes beyond small perturbations around a fixed point. However specific behaviors such as intermittent bursts of activity that are observed in biological (both neuronal and gene) networks are not intended to be captured by these equations, and would required detailed information on the specific system. The analysis of the stability of the system for all the actual and randomized networks is shown in Figure S16.

S13. Link degree and bifan degree hub removal

In order to further analyze the effect of the link degree hubs, which tend to be exclusively out- or in-link nodes, on the anti-ferromagnetism properties of the networks, we implemented an algorithm to construct surrogate networks where the hub structure was progressively dismantled. Following the same rationale, we also created similar surrogate networks based on the elimination of *bifan degree* hubs, i.e. nodes that contribute to the formation of many bifans..
The algorithms for link-hub and bifan-hub removal are as follows:

A. ***Link degree hub removal***

1. Find the nodes with maximum out-link and in-link degrees
2. Construct a first surrogate network, by reassigning all the links (i.e. out- and in-links) for both top out- and in-hubs randomly among all nodes.
3. Construct a second surrogate network by randomly reassigning the same number of links as in (1), by picking links at random, as opposed to links from the top two hubs.
4. Compute the magnetization for loops of size 3 and 4, for the first and second surrogate networks.
5. Repeat steps (1-4) until a predetermined number of link-hubs are removed.

B. ***Bifan hub removal***

1. Find the nodes with maximum out-bifan and in-bifan degrees. The out-bifan degree of a node is defined as the number of bifan motifs that include the node as source, computed as:
$$2B_k = \left[AA^T AA^T\right]_{kk} - \left(D_k^{(O)}\right)^2 - \left[A\left(D^{(I)} - \vec{1}\right)\right]_k$$
where A is the adjacency matrix, k is the node, $D^{(O,I)}$ the out- and in-link degree vectors, and $\vec{1}$ is a vector of 1's.
2. The in-bifan degree of a node is defined as the number of bifan motifs that include the node as asink, computed by exchanging A with A^T, and (O) with (I) in the equation above.

3. Construct a first surrogate network by randomly reassigning all the links (i.e. out- and in-links) for both top out- and in-bifan hubs among all nodes.
4. Construct a second surrogate network by randomly reassigning the same number of links as in (1), by picking links at random, as opposed to links from the bifan hubs.
5. Compute the magnetization for loops of size 3 and 4, for the first and second surrogate networks.
6. Repeat steps (1)-(5) until a predetermined number of bifan-hubs are removed.

C. ***Computation of magnetization for loops of size 3 and 4***

In all cases the neutral links of the networks were converted to directed links by assigning random directionality. For each network, the process was repeated to generate several realizations for computing population statistics. For simplicity and ease of computation, we only analyzed the effect of hub removal on the magnetization of loops of size 3 and 4, which can be done efficiently using linear algebra. For loops of size 3, the magnetization can be computed as:

$$M_3 = \left[3P + 1(Q - P)\right]/Q$$

where $Q = tr\left[\left(A + A^T\right)^3\right]/3$ is the total number of triangles, regardless of orientation, and $P = tr\left(A^3\right)$ is the total number of directed feedback loops. For loops of size 4, the magnetization is:

$$M_4 = \left[4Q + 2S\right]/P$$

where

$$P = \frac{1}{8}\left(tr\left[\left(A + A^T\right)^4\right] - 4\sum_k \binom{D_k}{2} - 2\sum_{i,j} A_{ij} \right)$$

$Q = tr\left[A^4\right]/4$, $S = tr\left[A^3 A^T\right]$, and $D = D^{(O)} + D^{(I)}$.