# Model of brain activation predicts the neural collective influence map of the brain

Flaviano Morone[a,1], Kevin Roth[a,b,1], Byungjoon Min[a,1], H. Eugene Stanley[c,1,2], and Hernán A. Makse[a,1,2]

[a]Levich Institute and Physics Department, City College of New York, New York, NY 10031; [b]Theoretical Physics, Eidgenössische Technische Hochschule Zürich, 8093 Zürich, Switzerland; and [c]Center for Polymer Studies and Physics Department, Boston University, Boston, MA 02215

Contributed by H. Eugene Stanley, December 28, 2016 (sent for review April 21, 2016; reviewed by Nuno M. A. Araújo and Lucilla de Arcangelis)

**Efficient complex systems have a modular structure, but modularity does not guarantee robustness, because efficiency also requires an ingenious interplay of the interacting modular components. The human brain is the elemental paradigm of an efficient robust modular system interconnected as a network of networks (NoN). Understanding the emergence of robustness in such modular architectures from the interconnections of its parts is a longstanding challenge that has concerned many scientists. Current models of dependencies in NoN inspired by the power grid express interactions among modules with fragile couplings that amplify even small shocks, thus preventing functionality. Therefore, we introduce a model of NoN to shape the pattern of brain activations to form a modular environment that is robust. The model predicts the map of neural collective influencers (NCIs) in the brain, through the optimization of the influence of the minimal set of essential nodes responsible for broadcasting information to the whole-brain NoN. Our results suggest intervention protocols to control brain activity by targeting influential neural nodes predicted by network theory.**

brain | collective influence | robustness | network of networks | optimal percolation

Experience reveals that the brain is composed of massively connected neural elements arranged in modules (1, 2) spatially distributed, yet highly integrated to form a system of network of networks (NoN) (3–9). These modules integrate in larger aggregates to ensure a high level of global communication efficiency within the overall brain network, while preserving an extraordinary robustness against malfunctioning (3–5).

The question of how these different modules integrate to preserve robustness and functionality is a central problem in systems science (3–5). The simplest modular model (2) would assign the same function to the connections inside the modules and across the modules. However, the existence of modularity gives rise to two types of connections of intrinsically different nature: the intermodular links and intramodular links (6, 9–11). Intramodular links define modules usually composed of clustered nodes that perform the same specific function, like, for instance, the visual cortex responsible for processing visual information. Besides having intralinks, nodes in a given module may have intermodular connections to control or modulate the activity of nodes in other spatially remote modules (3, 5, 6, 9, 12).

For example, in integrative sensory processing, the intermodular links mediate the bottom-up (or stimulus-driven) processes from lower-order areas (e.g., visual) to higher-order cortical ones, and top-down (or goal-directed) control from higher levels to lower ones (3, 5, 6, 12). Indeed, in studies of attention, the pattern of brain activation indicates that high-level regions in dorsal parietal and frontal cortex are involved in controlling low-level visuospatial areas forming a system of networks connected through intermodular control links (dorsal-frontoparietal NoN) (6, 12). The purpose of this work is to introduce a minimal model for a robust brain NoN made of such intramodule connections and intermodular controllers, which, by abstracting away complexity, will allow us to make falsifiable predictions about the location of the most influential nodes in the brain NoN. Targeting these neural collective influencers (NCIs) may help in designing intervention protocols to control brain activity prescribed by network theory (13, 14).

## Results

We consider a substrate NoN composed by two modules (Fig. 1*A*; below, we generalize to more modules). Every node $i$ has $k_i^{\text{in}}$ intramodular links to nodes in the same module and $k_i^{\text{out}}$ intermodular links to control other modules (for the sake of simplicity, we first consider the case $k_i^{\text{out}} \in \{0, 1\}$ for every node $i$; the general case $k_i^{\text{out}} \in \mathbb{N}_0$ will be treated later). Because controlling links connect two different modules, they are fundamentally different from intramodular ones: The latter encode only the information about if two nodes are connected or not inside a module, whereas the former carry the additional information about how nodes control each other in two different modules. We arrive to an important difference between both types of links that has been recognized in previous NoN models (10). An intermodular link between two nodes exists because of their mutual dependence across two distinct modules performing different functions. Therefore, it is reasonable that, for this intermodular link to be active, both nodes across the modules should be active. On the contrary, nodes inside a module connected only via intramodular links that do not participate in intermodular dependencies will be active independently on the other module's activity. The intralinks and interlinks are analogous to the strong and weak links defining hierarchical modules in the NoN in refs. 9 and 11.

### Significance

**Evidence suggests that the brain is arranged in functionally specialized modules to form a network of networks (NoN). Understanding how functionality emerges from the integration of such modular architectures is one of the greatest scientific challenges. We introduce a model of brain NoN, which is robust against random node failures, captures the integration of functionally specialized modules in the brain, and provides falsifiable predictions about the locations of the most influential nodes, called neural collective influencers, in the brain network—predictions that are impossible in existing but fragile models of interdependent NoN. If confirmed by experiment, our results may pave the way for applications of clinical interest.**

Author contributions: H.E.S. and H.A.M. designed research; F.M., K.R., B.M., and H.A.M. performed research; F.M., K.R., B.M., and H.A.M. analyzed data; and F.M., K.R., H.E.S., and H.A.M. wrote the paper.

Reviewers: N.M.A.A., Universidade de Lisboa; and L.d.A., Second University of Naples.

The authors declare no conflict of interest.

[1]F.M., K.R., B.M., H.E.S., and H.A.M. contributed equally to this work.

[2]To whom correspondence may be addressed. Email: hmakse@lev.ccny.cuny.edu or hes@bu.edu.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1620808114/-/DCSupplemental.

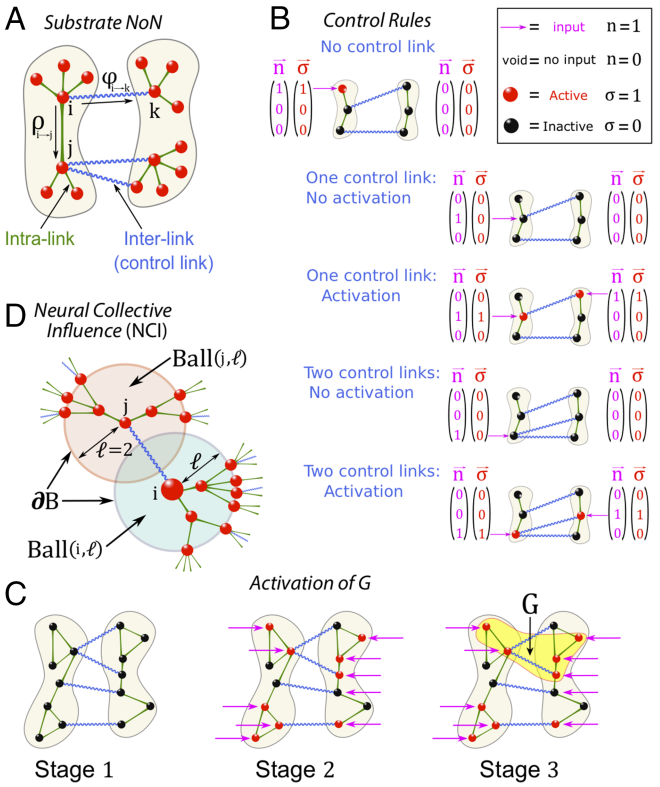


**Fig. 1.** Definition of NoN model. (*A*) Substrate NoN. Each node has $k_i^{\text{in}}$ intramodular links and $k_i^{\text{out}}$ intermodular control links. Nodes send information through two messages to their neighbors: a message $\rho_{i\to j}$ along the intralink and a message $\varphi_{i\to k}$ along the control link. (*B*) Control rule Eq. **3**. A node *i* in the substrate NoN may receive an external input $n_i = 1$, or not $n_i = 0$. If the node has no control link, it activates as soon as it receives the external input: $n_i = 1 = \sigma_i$. If it has 1 control link, it activates $\sigma_i = 1$ if and only if (iff) it receives the input, $n_i = 1$, and its neighbor at the edge of the interlink receives the input as well ($n_j = 1$). If it has 2 control links (or more) it activates ($\sigma_i = 1$) iff it receives the input and at least one node among its *j* neighbors at the edge of the interlink also receives an input $n_j = 1$, otherwise it does not activate ($\sigma_i = 0$). (*C*) Activation of the mutual giant component. Global communication in the NoN is measured through the largest active component, *G*, which is measured only with the active nodes $\sigma_i = 1$. We start with a NoN with no external input (all $n_i = 0$), then $G = 0$ (stage 1). Once an input is presented to the brain NoN (stage 2), nodes activate according to the rules in *B*, and the largest component of activated nodes defines *G* (stage 3), which is not necessarily equal to the sum of the individual giant components of the single networks. Note the crucial ingredient of the model (not shared by the model of ref. 10): Active nodes ($\sigma = 1$) may exist outside *G*, and they can have intermodular control links with other nodes outside the giant component. Thus, nodes can be active without being part of the giant component of their own network in contrast to the rules in ref. 10. (*D*) Collective influence. The collective influence of node *i* is determined by the sum of the degree of nodes in *G* on the surface of two balls of influence with radius $\ell$: ∂Ball(*i*, $\ell$) centered at *i*, and ∂Ball(*j*, $\ell$) centered at *j*, where *j* is a neighbor of *i* at the edge of an interlink having out-degree $k_j^{\text{out}} = 1$.

To elaborate on the mode of intermodular control, think of a node $i$ as a receiver of inputs external to the NoN, such as external sensory inputs to the primary visual cortex (Fig. 1*B* and *SI Appendix*). The input variable $n_i = 1, 0$ specifies whether $i$ receives the external input ($n_i = 1$) or not ($n_i = 0$). For example, in the visual system, $n_i = 1$ is the subset of nodes receiving inputs in the earlier stages in cortical sensory processing (6).

According to the discussion above, the input $n_i$ alone does not determine the activation/inactivation state of the node $i$, which we measure by the state variable $\sigma_i$ taking values $\sigma_i = 1$ if $i$ is activated, and $\sigma_i = 0$ if not. When $i$ has a control link with $j$ in another network, the state $\sigma_i$ is determined not only by the input $n_i$, but also by the input $n_j$ to $j$: node $i$ is activated $\sigma_i = 1$ only when both nodes receive the input ($n_i = 1$ and $n_j = 1$). On the contrary, when at least one of the $i, j$ nodes does not receive input ($n_i = 0$ or $n_j = 0$), node $i$ is shut down $\sigma_i = 0$. This top-down and bottom-up control between different modules is quantified by the following control rule, which acts as a logical AND between two controlling nodes (we consider $k_i^{\text{out}} = \{0, 1\}$; Fig. 1*B*):

$$\sigma_i = n_i\, n_j, \qquad \text{control rule for } k_i^{\text{out}} = 1. \qquad \textbf{[1]}$$

Because not all nodes participate in the control of other nodes, a certain fraction of them [determined by the degree distribution $P(k^{\text{out}})$] do not establish intermodular links with other nodes, $k_i^{\text{out}} = 0$. These nodes without control-links (Fig. 1*B*) activate as soon as they receive an external input, that is,

$$\sigma_i = n_i, \qquad \text{control rule for } k_i^{\text{out}} = 0. \qquad \textbf{[2]}$$

Generalization of the control rule to more than one control link per node can be done in different ways. Here, we consider that a node is activated ($\sigma_i = 1$) iff it receives the input $n_i = 1$ and at least one among the nodes $j$ in another module connected to $i$ via a control link receives also an input (i.e., $n_j = 1$). Otherwise $i$ is not activated (Fig. 1*B*). Mathematically:

$$\sigma_i \;=\; n_i \left[1 - \prod_{j \in \mathcal{F}(i)} (1 - n_j)\right], \qquad \text{general control rule} \qquad \textbf{[3]}$$

where $\mathcal{F}(i)$ is the set of $k_i^{\text{out}}$ nodes connected to $i$ via intermodular control links. In the following, we always refer to the general control model Eq. **3**, unless stated otherwise.

The distinction between $n_i$ and $\sigma_i$ models the initial sensory inputs ($n_i$) and the final state response of the brain ($\sigma_i$) to those stimuli from top-down and bottom-up influences (6). Thus, the final state of the brain network $\sigma_i$ encodes the brain's interpretation of the world by modulating external input $n_i$ via controls (Eq. **3**) from other cortical areas (Fig. 1*C*). We note that a general model should explain brain activation, even when no external input is applied to the NoN (e.g., in resting-state brain). This may be accounted for by a dynamical system that drives the NoN into stable attractors, which in resting state may need no external input anymore.

Apart from receiving inputs $n_i$ and controlling other nodes via Eq. **3**, active nodes can also broadcast information to the network. When all nodes are active, the information sent by a node can reach the whole brain NoN. If some nodes become inactive (i.e., $\sigma_i = 1 \to \sigma_i = 0$), the remaining active nodes group together in disjoint components of active nodes, such that information starting from an active node in one active component cannot reach another active node in a different active component. We quantify the global communication efficiency of the brain NoN with the size of the largest (giant) mutually connected active component $G$ made of active nodes $\sigma_i = 1$ (stage 3 in Fig. 1*C*) (9–11). By strict definition, $G$ could be (almost) the entire brain (e.g., a visual stimulus sets off emotional cues, memory areas, etc.). In what follows, we will restrict the NoN to specific systems of interest in the brain, like the visual or motor system, which are identifiable by fMRI methods for a particular single task.

Each configuration of active/inactive nodes $\vec{\sigma} = (\sigma_1, \ldots, \sigma_N)$ is associated with a specific working mode of the brain. The plethora of different functions dynamically executed by the brain (4–7) results in the moment-by-moment changes of the configuration $(\sigma_1, \ldots, \sigma_N)$, and thus in different values of $G$. The crux of the matter is that, for typical input configurations $\vec{n} = (n_1, \ldots, n_N)$ [i.e., the ones produced by the majority of the external (e.g., visual) inputs], $G$ has to be large enough for a global integration of information from distributed areas in the brain. In other words, the brain NoN has to remain globally activated

during the acquisition of different inputs, meaning that $G$ has to be robust, and the more robust, the more states the brain can achieve. Therefore, a model of a brain NoN must be able to capture such robustness.

In our statistical mechanics approach, being robust means that the brain should develop an extensive $G$ for typically sampled configurations of the external inputs. As a first approximation, we assume that these inputs are sampled from a flat (random) distribution of $\vec{n}$. Thus, we first study the robustness of the NoN across the configurations of states typically sampled by the brain. The problem then becomes a classical percolation study of the NoN (10) following the activation/inactivation rule of Eq. **3**. Having established our model in the normal brain under typical inputs, we will then move to disease states, which impede global communication by annihilating focal essential areas in $G$ (13, 14).

We calculate $G$ induced by typical random configurations of inputs $\vec{n}$ as a function of the fraction $q = 1 - \langle n \rangle$ of zero inputs [these zero inputs are analogous to removed nodes in classical percolation (9–11)], and we show that $G$ remains sizeable, even for high values of $q$, thus probing the robustness of the model NoN. At a critical value $q_{\text{rand}}$, we find the random percolation critical point $G(q_{\text{rand}}) = 0$ (9–11) separating a globally connected phase with nonzero $G\,(q < q_{\text{rand}}) > 0$ from a disconnected phase $G(q > q_{\text{rand}}) = 0$ composed of fragmented subextensive clusters with no giant component in the thermodynamic limit. The most robust NoN is tantamount to a system with no disconnected phase (i.e., with a large value of $q_{\text{rand}}$, ideally $q_{\text{rand}} = 1$). That is, the brain is robust if it can sustain a well-defined giant connected component for as many typical inputs as possible.

The dynamics of information flow in the NoN is defined as follows. Generally speaking, each node processes activity from neighboring nodes. Here, we abstract this coding process by considering that nodes receive information from other nodes via "messages" containing the information about their membership in $G$. Based on the information they receive, nodes broadcast further messages, until they eventually agree on who belongs to $G$ across the whole network. Because there are two types of links, we define two types of messages: $\rho_{i\to j} \in \{0, 1\}$ running along an intramodular link, and $\varphi_{i\to j} \in \{0, 1\}$ running along an intermodular control link, where $\{0, 1\}$ represents a {no, yes} "*I belong to G*" message, respectively (Fig. 1*A*).

In this view, the existence of an extensive giant mutually connected component across the NoN, $G > 0$, expresses a percolation phase produced by the binding of activation patterns across different modules in a distributed emergent global system. Under this interpretation, perception is not the responsibility of any particular cortical area, but is an emergent critical property of the percolation of memberships interchanged across all members of $G$ (15). The percolation critical point $q_{\text{rand}}$ can be interpreted as the transition between a phase of global perception $G > 0$ for $q < q_{\text{rand}}$ and a null perception phase characterized by nonextensive disconnected components and the concomitant $G = 0$ for $q > q_{\text{rand}}$.

The equations governing the information flow in the brain NoN follow the updating rules of the membership messages according to (analytical details in *SI Appendix*):

$$\rho_{i\to j} = \sigma_i\Big[1 - \prod_{k\in\mathcal{S}(i)\setminus j}(1-\rho_{k\to i})\prod_{l\in\mathcal{F}(i)}(1-\varphi_{l\to i})\Big],$$
$$\varphi_{i\to j} = \sigma_i\Big[1 - \prod_{k\in\mathcal{S}(i)}(1-\rho_{k\to i})\prod_{l\in\mathcal{F}(i)\setminus j}(1-\varphi_{l\to i})\Big], \quad [4]$$

where $S(i)\setminus j$ is the set of $k_i^{\text{in}} - 1$ neighbors of node $i$ in the same module, except $j$. Eq. **4** indicates, for instance, that a positive membership message $\rho_{i\to j} = 1$ is transmitted from node $i \to$ node $j$ in the same module (analogously, $\varphi_{i\to j}$ transmits messages to the other module) if node $i$ is active $\sigma_i = 1$ and if it receives at least one positive message from either a node $k$ in the same module $\rho_{k\to i} = 1$ or a node $l$ in the other module $\varphi_{l\to i} = 1$. The logical OR is important; it is the basis for such a robust R-NoN brain model of activation as elaborated below.

To compute $G$, it is sufficient to know for each node $i$ whether it is or not a member of $G$, which is encoded in the quantity $\rho_i \in \{0, 1\}$ representing the probability to belong to $G = \langle \rho_i \rangle$:

$$\rho_i = \sigma_i\Big[1 - \prod_{k\in\mathcal{S}(i)}(1-\rho_{k\to i})\prod_{l\in\mathcal{F}(i)}(1-\varphi_{l\to i})\Big]. \quad [5]$$

Here, we arrive to an important point (illustrated in Fig. 1*C*), which ultimately explains the robustness of our brain NoN: In our model, a node can be active ($\sigma_i = 1$) even if it does not belong to the giant mutually connected active component $G$, thus preventing catastrophic cascading effects. This feature of the brain model is supported by neuroanatomical correlates: The brain responds reasonably well to injuries, for instance, to areas such as the arcuate fasciculus (the white matter tract that connects the two most important language areas—Broca's and Wernicke's areas). This property is the main difference between our model and previous NoN models (10) describing catastrophic collapse in power-grids (16), as discussed next.

**Universality Classes of NoN.** In the model of ref. 10, a node can be active only if it belongs to the giant component in its own network. Thus, in this model, the active/inactive state of a node is controlled by the whole global giant component $\rho_i$, rather than the local state variable $\sigma_i$, Eq. **3**, as in our model. This means that, in ref. 10, the state of a node is actually controlled by the whole network [i.e., intermodular controls (therein called dependencies) carry the weight of the extensive giant component]. Analogously, the NoN cannot be built from the $G = 0$ phase, because it would require the existence of extensive components for each network. For this reason, the resulting NoN (10) is fragile; a single inactivation of a node can lead to catastrophic collapse of the whole active giant component [which, we note, can be avoided by strong correlations between the hubs of different networks (9)]. Conversely, the model of Eq. **3** allows nodes to be active, even if they do not belong to $G$ (i.e., when they belong to nonextensive disconnected components). These small components become crucial to build the $G > 0$ phase from the $G = 0$ phase by adding interlinks to nonextensive components.

Indeed, the model of ref. 10 was proposed to capture the fragility of certain manmade infrastructures, such as the catastrophic collapse of power grids (e.g., the Northeast U.S. blackout of 2003, which allegedly started in a single power-line failure as modeled in ref. 10). The equation to compute $G$ in this catastrophic C-NoN model reads:

$$\rho_i = \sigma_i\Big[1 - \prod_{k\in\mathcal{S}(i)}(1-\rho_{k\to i})\Big]\Big[1 - \prod_{k\in\mathcal{F}(i)}(1-\varphi_{k\to i})\Big]. \quad [6]$$

We note that Eq. **6** differs from R-NoN Eq. **5** in that the logical OR has been replaced by the logical AND for message passing in C-NoN.

A third possible model for NoN is the modular model (2) mentioned in the introduction, which assumes no difference between intralinks and interlinks as studied in ref. 17. In this model, there are no control links; therefore, nodes cannot control each other, and the state equals the input: $\sigma_i = n_i$. This model is described by using only the intralink messages, $\rho_{i\to j}$, corresponding to a single network structure, albeit with modularity (2), and $\rho_i$ is simple, given by (no special messages between modules):

$$\rho_i = n_i\Big[1 - \prod_{k\in\mathcal{S}(i)}(1-\rho_{k\to i})\Big]. \quad [7]$$

We thus arrive to three different universality classes of NoN—R-, C-NoN, and modular single network—according to the three models given by Eqs. **5**–**7**, respectively, which are defined according to which variable controls the state of node $i$ ($\sigma_i$, $\rho_i$, $n_i$) (Table 1). Among the three universality classes, only R-NoN is robust with the functionality of control across modules via top-down and bottom-up influences.

**Robustness of the Brain NoN to Typical Inputs.** We compute $G(q)$ from Eq. **5** when we present different typical random inputs $n_i$ and show that the obtained percolation threshold $q_{\text{rand}}$ is close to 1. The results are first tested on synthetic NoN made of Erdös–Rényi (ER) and scale-free (SF) random graphs (1).

Results in Fig. 2*A* show that our model indeed defines a robust R-NoN characterized by large $q_{\text{rand}}$. Additionally, Fig. 2*B* compares model R-NoN Eq. **5** with the catastrophic C-NoN universality class Eq. **6**, showing that these two models capture two different phenomena, the former robust with larger $q_{\text{rand}}$ and second-order phase transitions, and the latter catastrophic with smaller $q_{\text{rand}}$ with first-order abrupt transitions.

**Response to Rare Events: NCIs.** Having investigated the behavior of the model under typical inputs, we now study the response of the brain NoN to rare events targeting a set of NCIs. These are rare inputs: An optimal (minimal) set of nodes that when they are shut down ($n_i = 0$) disintegrates the giant component to $G = 0$ using the smallest possible fraction of nodes, $q_{\text{infl}}$. This is the process of optimal percolation (rather than classical random percolation treated above) as defined in ref. 18. The malfunction of these neural influencers could be associated with pathological states of the brain arising from interruption of global communication in the network structure, such as depression or Alzheimer's disease. The underlying conjecture is that these influencers could be responsible for neurological disorders (13, 14). At the same time, activating this minimal set of neural influencers, ($n_i = 1$, $\sigma_i = 1$), would optimally broadcast the information to the entire network (19). Thus, these neural influencers are also the minimal set of nodes that provide integration of global activity in the NoN (15).

Finding this minimal set is a NP-hard combinatorial optimization problem (19). Here, we follow ref. 18, which developed the theory of optimal percolation for a system with a single network and solve the problem in a NoN. As opposed to random percolation that identifies $q_{\text{rand}}$, optimal percolation identifies the minimal fraction of influencers $q_{\text{infl}}$ that, if removed, optimally fragment the giant connected component [i.e., with minimal removals ($n_i = 0$)]. We note that these neural influencers are statistically rare (i.e., they cannot be obtained by random sampling $\vec{n}$).

The mapping to optimal percolation (18) allows us to find brain influencers under the approximation of a sparse graph by minimizing the largest eigenvalue $\lambda(q, \vec{n})$ of a modified nonbacktracking (NB) matrix (20) $\mathcal{M}_{\rho\varphi} \equiv (\partial\rho_{i\to j}/\partial\varphi_{k\to\ell})_{\rho=\varphi=0}$ of the NoN over all configurations of inputs $\vec{n}$ having a fraction $q$ of zero inputs (analytical details in *SI Appendix*). The NB matrix $\hat{\mathcal{M}}$ controls the stability of the solution of the broken phase $G = 0$. This solution becomes unstable (i.e., $G$ becomes nonzero) when the largest eigenvalue is 1. The minimal set of influencers $\vec{n}_{\text{infl}}$ and their fraction $q_{\text{infl}}$ are then found by solving: $\lambda(q_{\text{infl}}, \vec{n}_{\text{infl}}) = \min_{\vec{n}} \lambda(q_{\text{infl}}, \vec{n}) = 1$.

The eigenvalue $\lambda(\vec{n})$ can be efficiently minimized by progressively removing the input ($n_i = 1 \to n_i = 0$) from the nodes with the highest Collective Influence index $\text{CI}_\ell(i)$ (detailed derivation in *SI Appendix*) given by ($z_i \equiv k_i^{\text{in}} + k_i^{\text{out}} - 1$):

$$\text{CI}_\ell(i) = z_i \sum_{j\in\partial\text{Ball}(i,\ell)} z_j + \sum_{\substack{j\in\mathcal{F}(i):\\ k_j^{\text{out}}=1}} z_j \sum_{m\in\partial\text{Ball}(j,\ell)} z_m. \quad [8]$$

The collective influence $\text{CI}_\ell(i)$ of node $i$ is determined by two factors (Fig. 1*D*). The first one is a node-centric contribution, given by the first term in Eq. **8**, where $\text{Ball}(i, \ell)$ is the set of nodes inside a ball of radius $\ell > 0$ ($\ell$ is the distance of the shortest path between two nodes), centered on node $i$, and $\partial\text{Ball}(i, \ell)$ its frontier. This ball is grown from the central node $i$ by following both intralinks and interlinks, and thus may invade different networks in the NoN. The second factor is a node-eccentric

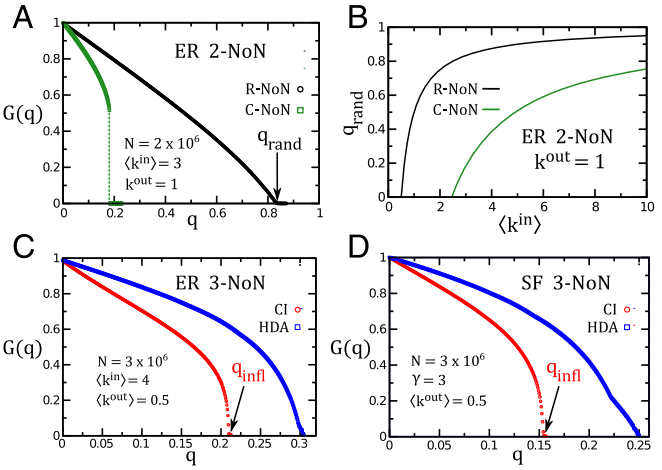


**Fig. 2.** Robustness and NCI in NoN. (*A*) Robustness of NoN under typical random inputs. Size of the largest active component $G(q)$ for typically sampled inputs $\vec{n}$ for ER 2-NoN (meaning a NoN made of two ER networks) for the R- and C-NoN universality classes ($k^{\text{out}} = 1$ for all nodes, one-to-one control links, total size $N = 2 \times 10^6$). The large value of $q_{\text{rand}}$ in R-NoN compared with C-NoN confirms the robustness of the former. The transition separating the phases $G = 0$ and $G > 0$ is 2$^{\text{nd}}$-order in R-NoN and 1$^{\text{st}}$-order in C-NoN, reinforcing the fundamental difference (robust vs. fragile) of these two universality classes (errors are SEM over 10 realizations). (*B*) Phase diagram for R- and C-NoN. Behavior of $q_{\text{rand}}$ as a function of the average $\langle k^{\text{in}}\rangle$ for the ER 2-NoN in *A*, where each node has $k^{\text{out}} = 1$. Here, $q_{\text{rand}}$ is the fraction of nodes with zero inputs in one network (nodes in the other network have all nonzero inputs). The difference in $q_{\text{rand}}$ between R- and C-NoN ranges from 20% for $\langle k^{\text{in}}\rangle = 10$ to 80% for $\langle k^{\text{in}}\rangle \sim 2.5$. Analytically, we find for R-NoN with $k^{\text{out}} = 1$, $q_{\text{rand}} = 1 - 1/(2\langle k^{\text{in}}\rangle)$. (*C*) Rare inputs and NCI in ER 3-NoN. Size of $G(q)$ as a function of the untargeted ($n_i = 0$) nodes $q$ for a NoN of three ER networks (total size $N = 3 \times 10^6$). Each network has $10^6$ nodes, $\langle k^{\text{in}}\rangle = 4.0$ and $\langle k^{\text{out}}\rangle = 0.5$. We show the CI optimization (red circles, $\ell = 4$) and the high-degree adaptive (HDA) heuristic (blue squares; removal by highest $k^{\text{in}}$) (21). The arrow marks the position of the minimal fraction of influencers $q_{\text{infl}}$, which is smaller than the HDA centrality (errors are SEM over 10 realizations). Other heuristic centralities perform worse than HDA. (*D*) Rare inputs and NCI in SF 3-NoN. $G(q)$ for a NoN with three SF networks (total size $N = 3 \times 10^6$). Each network is SF with $10^6$ nodes, minimum and maximum degree $k_{\text{min}}^{\text{in}} = 2$ and $k_{\text{max}}^{\text{in}} = 10^3$, and power-law exponent $\gamma = 3$. The node out-degree is Poisson-distributed with average $\langle k^{\text{out}}\rangle = 0.5$ (errors are SEM over 10 realizations). The difference between CI ($\ell = 3$) and HDA is shown; HDA fails to identify 40% of influencers.

**Table 1. Universality classes of NoN**

| Universality class | State control | Robust | Control functionality |
|---|---|---|---|
| Brain robust R-NoN Eq. **5** | $\sigma_i$ | Yes | Yes |
| Power-grid catastrophic C-NoN Eq. **6** | $\rho_i$ | No | Yes |
| Modular single network Eq. **7** | $n_i$ | Yes | No |

contribution, given by the second term in Eq. **8**, where the sum runs over all nodes $j$ connected to $i$ by an interlink which have out-degree equal to one $k_j^{\text{out}} = 1$ (this means that nodes $j$ have no other interlinks except to node $i$). The contribution of each of these $j$ nodes is given by growing another ball $\text{Ball}(j, \ell)$ around them. This last contribution is absent in the single network case (18), and thus, it is a genuine new feature of the brain NoN.

The NCI are formally defined as the nodes in the minimal set up to $q_{\text{infl}}$. To identify them, we start with all $n_i = 1$ and $\sigma_i = 1$, and we progressively remove one by one the inputs (setting $n_i = 1 \to n_i = 0$) to the nodes having the largest $\text{CI}_\ell(i)$ value if they are active $\sigma_i = 1$. At each step, the $\text{CI}_\ell(i)$ values are recomputed, and the algorithm (described in detail in *SI Appendix*) stops when $G = 0$ where the NCI set is identified.

We first test our predictions on influencers using synthetically generated ER-NoN and SF-NoN. Fig. 2*C* and *D* show the optimality (smaller $q_{\text{infl}}$) of our predicted set of influencers in comparison with the high-degree centrality (21), a heuristic commonly used in graph analysis of pathological brain networks (14). The theory allows us to predict the neural collective influence map (NCI-map) of the brain as explained next.

**NCI-Map of the NoN.** We apply our model to a paradigmatic case of stimulus-driven attention (9, 11, 22). The experiment consists of a dual visual–auditory task performed by 16 subjects (*SI Appendix*). Each subject received simultaneously a visual stimulus and an auditory pitch, to which the subject has to respond with the right hand if a number was larger than a reference and with the left hand if a tone was of high frequency.

The rationale to choose this experiment, where stimuli are received simultaneously, is that this imposes to select an appropriate response order with consequent deployment of high-level control modules in the brain (22). This effect emphasizes the role of top-down control of intermodular links that is the main effect we are trying to capture in our model.

The brain NoN was inferred from the brain activity recorded through functional magnetic resonance imaging (fMRI). Nodes in the NoN represent fMRI voxels whose size is given by the normalized spatial resolution of the fMRI scan $2 \times 2 \times 2$ mm$^3$. Pairwise cross-correlation between the BOLD signals of two nodes represents only indirect correlations (known as the functional connectivity network) capturing the weighted sum of all possible direct interactions between two nodes that could arise from the underlying unknown structural network and other interactions modulating the activity of neurons (7). To construct the brain NoN, we infer the strength of these interactions between nodes by using machine learning maximum-entropy methods (23–25), where we maximize the likelihood of the interactions between nodes given the observed pattern of fMRI cross-correlations (full details in *SI Appendix*). The resulting NoN is shown in Fig. 3*A* and *B*, which is then used to identify the NCI in the brain network activated for this particular task.

In all subjects, we observe the following (Fig. 3 *A* and *B*): (*i*) A network partially covering the anterior cingulate (AC) region, recruited for decision making and therefore processing top-down and bottom-up control; (*ii*) a network covering the medial part of the posterior parietal cortex (PPC), which receives somatosensory inputs and sends the output to areas of the frontal motor cortex to control particular movements of the arms; and (*iii*) a network covering the medial part of the posterior occipital cortex (area V1/V2), along the calcarine fissure, which is responsible for processing visual information at lower input levels (an additional auditory network was also observed; *SI Appendix*).

We apply our theory to the AC-PPC-V1/V2 3-NoN to first test the robustness under typical inputs and then obtain the NCI (rare inputs). Indeed, the obtained brain 3-NoN is very robust to typical inputs, as shown by the large (close to one) $q_{\text{rand}} \approx 0.9$ in

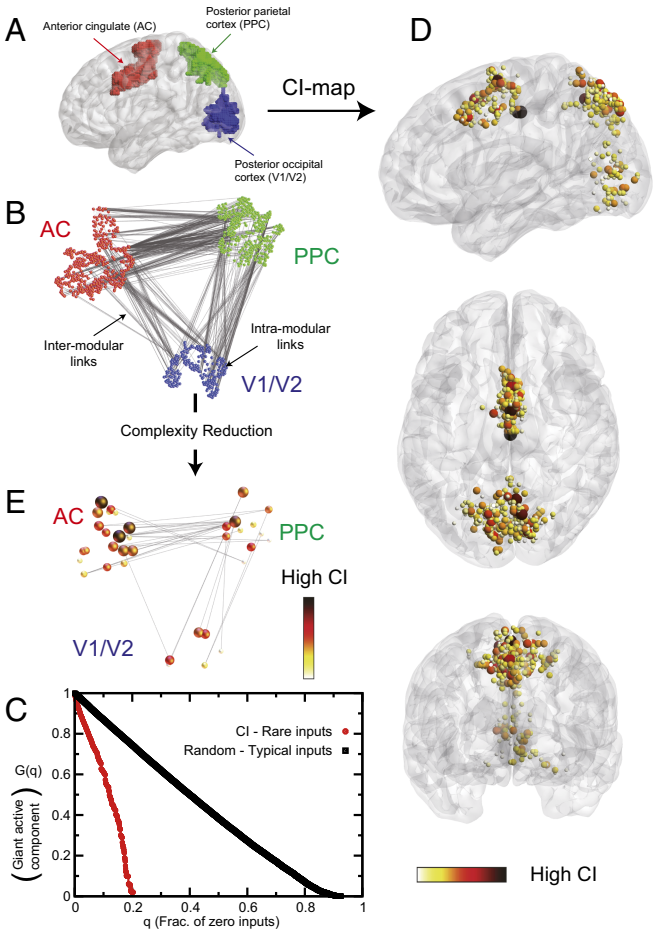


**Fig. 3.** Brain-NoN. (*A*) The 3-NoN in dual-task fMRI experiment. Shown is the spatial location of the three main networks for a typical subject (as opposed to averaging over all subjects as in *D*) showing the anterior cingulate (AC; red), posterior parietal cortex (PPC; green), and posterior occipital visual areas V1/V2 (blue). This 3-NoN structure appears consistently for all 16 subjects. Nodes in the NoN represent voxels in the fMRI BOLD signal of normalized size $2 \times 2 \times 2$ mm$^3$. (*B*) Topology of the 3-NoN. Same as *A*, but in the network representation with interlinks in gray. Number of nodes in NoN is $N = 1{,}134$, $\langle k^{\text{in}} \rangle = 3.2$, and $\langle k^{\text{out}} \rangle = 2.5$. (*C*) Robustness and NCI. Size of the largest active cluster $G(q)$ as a function of the untargeted ($n_i = 0$) nodes $q$ after CI optimization (red curve; $\ell = 3$) and after typical random states (black; random percolation). (*D*) NCI-map of the human brain averaged over 16 subjects. The color code ranges from 0 to 5.2 and represents the number of subjects in which a node appears in the ranked NCI set (*SI Appendix*). High-CI influential regions are located mainly in the AC module for processing top-down control, whereas the influential nodes are rarely located in the lower-level V1/V2 region. The PPC region contains a portion of influential nodes closer to AC. (*E*) Complexity reduction to top NCI nodes. Controlling links between different networks are mainly mediated by top influencers.

Fig. 3*C* (black curve). Conversely, the theory is able to localize the minimal set of NCI with $q_{\text{infl}} \approx 0.2$ (Fig. 3*C*, red curve). Using these influential nodes, we construct the NCI-map averaging over all subjects. The emerging NCI-map averaged over the 16 subjects is portrayed in Fig. 3*D* (details in *SI Appendix*). We find that the main influence region (high CI) is located mainly in the AC module as expected, because AC is a central station of top-down control. The areas of high influence also extends to a portion of the PPC involved in both top-down and bottom-up control, and it is less prominent in the V1/V2 areas, which are mostly involved in processing input information and bottom-up interactions. Therefore, the NCI-map of the brain suggests that control is

deployed from the higher level module (AC) toward certain strategic locations in the lower ones (PPC–V1/V2), and these locations can be predicted by network theory. The complexity reduction obtained by coarse-graining the whole NoN to the top NCI in Fig. 3*E* highlights the predicted strategic areas in the brain.

## Discussion

We present a minimal model of a robust NoN to describe the integration of brain modules via control interconnections. The key point of the model is that a node can be active, even if it does not belong to the giant mutually connected active component so that cascades are not fatal. Although our model is expressed in abstracto by logic relations, it is able to make falsifiable predictions (e.g., the location of the most influential neural nodes involved in information processing in the brain). If confirmed experimentally, our results may have applications of clinical interest, in that they may help to design therapeutic protocols to handle pathological network conditions and to retune diseased network dynamics in specific neurological disorders with interventions targeted to the activity of the influential nodes predicted by network theory. On the theoretical side, further extensions of our model are also possible. For instance, the model could be enriched by incorporating temporal dependence of brain activation, which is relevant for the theoretical description of synaptic plasticity (26).

**ACKNOWLEDGMENTS.** We thank S. Canals, S. Havlin, and L. Parra for discussions and M. Sigman for providing the data. This work was supported by National Science Foundation (NSF) Grants PHY-1305476 and IIS-1515022; NIH-National Institute of Biomedical Imaging and Bioengineering Grant 1R01EB022720; NIH-National Cancer Institute U54CA137788/U54CA132378; and Army Research Laboratory (ARL) Grant W911NF-09-2-0053 (ARL Network Science Collaborative Technology Alliance). The Boston University work was supported by NSF Grants PHY-1505000, CMMI-1125290, and CHE-1213217, and by Defense Threat Reduction Agency Grant HDTRA1-14-1-0017 and Department of Energy Contract DE-AC07-05Id14517.

1. Caldarelli G, Vespignani A (2007) Large Scale Structure and Dynamics of Complex Networks: From Information Technology to Finance and Natural Science (World Scientific, Singapore).
2. Newman MEJ (2006) Modularity and community structure in networks. *Proc Natl Acad Sci USA* 103:8577–8582.
3. Tononi G, Sporns O, Edelman GM (1994) A measure for brain complexity: Relating functional segregation and integration in the nervous system. *Proc Natl Acad Sci USA* 91:5033–5037.
4. Treisman A (1996) The binding problem. *Curr Opin Neurobiol* 6:171–178.
5. Dehaene S, Naccache L (2001) Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework. *Cognition* 79:1–37.
6. Corbetta M, Shulman GL (2002) Control of goal-directed and stimulus-driven attention in the brain. *Nat Rev Neurosci* 3:201–215.
7. Bullmore E, Sporns O (2009) Complex brain networks: Graph theoretical analysis of structural and functional systems. *Nat Rev Neurosci* 10:186–198.
8. Russo R, Herrmann HJ, de Arcangelis L (2014) Brain modularity controls the critical behavior of spontaneous activity. *Sci Rep* 4:4312.
9. Reis SDS, et al. (2014) Avoiding catastrophic failure in correlated network of networks. *Nat Phys* 10:762–767.
10. Buldyrev SV, Parshani R, Paul G, Stanley HE, Havlin S (2010) Catastrophic cascade of failures in interdependent networks. *Nature* 464:1025–1028.
11. Gallos LK, Makse HA, Sigman M (2012) A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proc Natl Acad Sci USA* 109:2825–2830.
12. Gilbert CD, Sigman M (2007) Brain states: Top-down influences in sensory processing. *Neuron* 54:677–696.
13. Stam CJ (2014) Modern network science of neurological disorders. *Nat Rev Neurosci* 15:683–695.
14. van den Heuvel MP, Sporns O (2013) Network hubs in the human brain. *Trends Cogn Sci* 17:683–696.
15. Crick F, Koch C (2003) A framework for consciousness. *Nat Neurosci* 6:119–126.
16. Rosato V, et al. (2008) Modeling interdependent infrastructures using interacting dynamical models. *Int J Crit Infrastruct* 4:63–79.
17. Leicht EA, D'Souza RM (2009) Percolation on interacting networks. arXiv:0907.0894.
18. Morone F, Makse HA (2015) Influence maximization in complex networks through optimal percolation. *Nature* 524:65–68.
19. Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. Proceedings of the 9th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (Association for Computing Machinery, New York), pp 137–143.
20. Hashimoto K (1989) Zeta functions of finite graphs and representations of p-adic groups. *Adv Stud Pure Math* 15:211–280.
21. Albert R, Jeong H, Barabási A-L (2000) Error and attack tolerance of complex networks. *Nature* 406:378–382.
22. Sigman M, Dehaene S (2008) Brain mechanisms of serial and parallel processing during dual-task performance. *J Neurosci* 28:7585–7598.
23. Schneidman E, Berry MJ, Segev R, Bialek W (2006) Weak pairwise correlations imply strongly correlated network states in a neural population. *Nature* 440:1007–1012.
24. Sarkar S, Chawla S, Xu D (2015) On inferring structural connectivity from brain functional-MRI data. arXiv:1502.06659.
25. Robinson PA, Sarkar S, Pandejee GM, Henderson J (2014) Determination of effective brain connectivity from functional connectivity with application to resting state connectivities. *Phys Rev E Stat Nonlin Soft Matter Phys* 90:012707.
26. Min B, et al. (2017) Finding essential nodes for integration in the brain using network optimization theory. Available at www-levich.engr.ccny.cuny.edu/hmakse/ltp23.pdf. Accessed March 17, 2017.