Creative destruction: collisions and redundancy generate emergent sparseness on the mammal connectome

Yan Hao (hao@hws.edu)
Department of Mathematics and Computer Science, Hobart & William Smith Colleges
Geneva, NY 14456 USA

Daniel Graham (graham@hws.edu)
Department of Psychology, Hobart & William Smith Colleges
Geneva, NY 14456 USA

ABSTRACT

Collision dynamics in brain network communication have been little studied. We describe a novel interaction that shows how nonlinear collision rules can result in efficient activity dynamics on simulated mammal brain networks. We tested the effects of collisions in "information spreading" (IS) models in comparison to standard random walk (RW) models. Simulations employ synchronous agents on tracer-based mesoscale mammal connectomes at a range of signal loads. We find that RW routing models have high average activity, which increases substantially with load. Activity in RW models is densely distributed over nodes: a substantial fraction are highly active in a given time window, and this fraction increases with load. Surprisingly, while IS routing models make many more attempts to pass signals, they show lower net activity due to collisions compared to RW. Activity in IS increases relatively little over a wide range of loads. In addition, IS models have greater sparseness (which decreases slowly with load) compared to RW models. Results hold on two networks of the monkey cortex and one of the mouse whole-brain. We also find evidence that activity is lower and more sparse for empirical networks compared to degree-matched randomized networks in IS, suggesting network topology supports IS routing.

Keywords: connectome; communication systems; collisions; redundancy; routing; sparseness

Introduction

With the advent of detailed connectomes in several mammals, researchers are increasingly aiming to "animate" these networks in order to understand patterns of whole-brain dynamics (a research area termed *dynomics*; Kopell et al., 2014). At the level of mesoscale mammal brain network structure, comprising connectivity of tens to hundreds of brain areas over thousands to hundreds of thousands of connections, activity on the network will not necessarily be linearly related to single-neuron spiking dynamics. There are many varieties of nonlinear and emergent interactions that could be at play, but these are difficult to study with linear models. One approach is to consider nodes as communicators that attempt to pass signals or messages to other nodes according to routing rules (Poggio, 1984; Hahn et al., 2018). Effective and robust routing strategies are especially important on the connectome because the topology is such that there are paths of just a few synapses between any node and practically any other node. Given the need to exchange multimodal information in real time—without changing network topology—the brain would appear to require a dynamic routing scheme that achieves efficient and reliable communication.

A given topology can support a range of routing strategies, giving rise to different observable dynamics (see, e.g., Friston, 2011). To begin to understand what routing strategy is in use in the brain, one can look to design considerations in engineered systems (Graham and Rockmore, 2010; Graham 2014; 2017; Navlakha et al., 2015; 2017; Fornito et al., 2016). A fundamental challenge for any large-scale communication system is the management of collisions. Collisions are the price paid for the ability to route signals selectively and dynamically to many possible destinations. Collision dynamics are emergent: they are a nonlinear effect of node dynamics, topology, and current traffic. All large-scale engineered communication systems have a means of managing collisions, through redress, arbitration, and other strategies (see e.g., Kleinrock, 1976; Mišić and Mišić, 2014).

In the brain, collisions can occur when signals from different areas converge on the same target at the same time. They also occur at the level of individual neurons, where they are traditionally conceived as a summation. Here, we reframe the concurrent arrival of signals as a collision, rather than as a summation. We discuss this reframing and the potential neural mechanisms for mediating collisions below. But regardless of the mechanisms involved, we argue that it is in the interest of mammalian brain networks to establish dynamic interactions that manage collisions successfully, especially given the a priori likelihood of collisions on small world-like networks (compared to, for example, lattice networks). As we discuss later, systems with collision rules such as thresholds and summations tend to produce undesirable behavior such as dying-off or overload.

All but a handful of attempts to model brain network dynamics to date have ignored the possibility of collisions. Random walk models (e.g., Abdelnour et al., 2014) which route signals to randomly chosen outgoing edges (either with uniform probability, in proportion to edge weight, or in other ways) have only rarely considered collisions. An exception is Mišić et al. (2014a,b). In this pair of studies, buffers (node memory allocations) were employed to manage collisions in a random walk context. In this scheme, colliding signals at the inputs of a node are lined up in a queue and stored in node memory until they can be directed to the designated output edge. Buffers are ubiquitous on the Internet and may be useful in brains. Neurobiological mechanisms for buffering have been proposed, but remain hypothetical (see, e.g., Goldman-Rakic, 1996; Graham, 2014; Funahashi, 2015).

Shortest path models (e.g., Bullmore and Sporns, 2009; Rubinov and Sporns, 2010) also typically ignore collisions. Invoking the logic of evolved optimality, shortest path routing models offer an alternative approach compared to random walks. Shortest path strategies also have parallels in engineered systems like the Internet (e.g., OSPF: open shortest-path first routing, see, e.g., Sosnovich et al., 2017). However, shortest path routing schemes must still manage collisions. Even when short path models of the dynamic mammal connectome include small buffers, they show substantial message loss due to congestion (Graham and Hao, 2018). In any case, shortest path routing requires biologically implausible mechanisms: to send messages on the shortest path successfully, the system needs global knowledge of network architecture, as well as of current network traffic. It is doubtful the brain could accomplish this (Seguin et al., 2018; but see Mišić et al., 2015).

Collisions need not incapacitate a communication network. They may be useful in establishing stable and robust dynamics. In particular, collisions that are destructive (i.e., that delete all colliding messages) could—almost paradoxically—help promote system efficiency. First, destructive collisions may help ensure a low average level of energy use, which is imperative given that the dynamic connectome of mammals has strong metabolic constraints. They may also promote homeostasis, which appears desirable for maintaining relatively constant activity in the face of changing system demands (e.g., Aeschbach et al., 1997; Turrigiano, 1999). However, the brain cannot achieve efficiency simply by using as little energy

as possible (Poldrack, 2015). The brain must also operate in sparse fashion, in part because the environment is sparse. It is important to note that sparse activity, while generally conferring low average energy usage, does not necessarily imply minimal energy use. Instead, it involves a non-Gaussian distribution of activity across units, in particular a distribution with a strong peak near zero activity (most units are "off" at a given time) and heavy tails (a small number of units are likely to be highly active). In consequence, only a small fraction of units—10% or less—in a given part of the network can be highly active at a given time (Levy and Baxter 1997; Lennie, 2000; Attwell and Laughlin, 2003); this is known as population sparseness. In addition, a neuron can only be highly active over a small fraction of its lifetime; this is known as lifetime sparseness (see Willmore and Tolhurst, 2001; Graham and Field, 2006). Sparseness in the visual system, for example, is thought to be in part a result of sparse physical environments (Field, 1994; Olshausen and Field 1996; Bell and Sejnowski, 1997). We suggest here that sparseness could also be an emergent property of brain network communication given brain network topology and nonlinear signal interactions.

If collisions are destructive, it may be necessary for the brain to adopt strategies to generate redundancy in order to promote reliability. One solution is to produce multiple copies of a message at each node. Based on this idea, we here introduce an *information spreading* (IS) model of brain network communication to investigate the effects of nonlinear collision rules. We compare this model to a standard random walk (RW) model.

The present work shows that sparseness can emerge as a feature of plausible routing strategies on the mammal connectome when nonlinear collisions are accounted for. We show that low activity and high sparseness across a range of loads are emergent properties of network-wide communication under an IS model. Performance of the IS model contrasts with RW models, which we show exhibit high and increasing energy use with increasing signal loads, and low and decreasing sparseness across loads. To investigate whether empirical topology in the mammal connectome supports sparse, efficient activity patterns, we also compare dynamics of the anatomical network to randomized networks with matched degree distributions.

Methods

Overview

Using a Markovian agent-based model on monkey and mouse tracer-based connectomes, we implemented an information spreading (IS) model wherein nodes pass copies of incoming signals to all output edges. Simulations utilized a nonlinear collision rule whereby all colliding messages are destroyed. We compared the IS model to a standard random walk (RW) model with the same collision rule. In RW, a single message is passed from a node to a randomly chosen outgoing edge.

Connectomes

We utilized three mesoscale mammal connectomes: two of the macaque monkey (Markov et al., 2014, termed '*monkey1*' in the present study as shorthand; and the CoCoMac database, see Bakker et al. 2012, termed '*monkey2*'), and one for the mouse (Oh et al., 2014, termed '*mouse*'). All edges are directed and have a weight of 1. The *monkey1* connectome comprises 91 cortical nodes and 1615 edges, primarily in the visual system (ipsilateral). Of these nodes, 29 have in- and out-degree > 0, while the remainder are source nodes (in-degree = 0). For the *monkey2* network, we used only nodes that had an in-degree and an

out-degree of at least 1. This set comprised 184 nodes corresponding to cortical regions and 5270 edges. The *mouse* connectome spans the entire mouse brain (ipsilateral), comprising 213 nodes and 16954 edges (all have in- and out-degree > 0). Adjacency matrices for the data sets are shown in Figure 1.

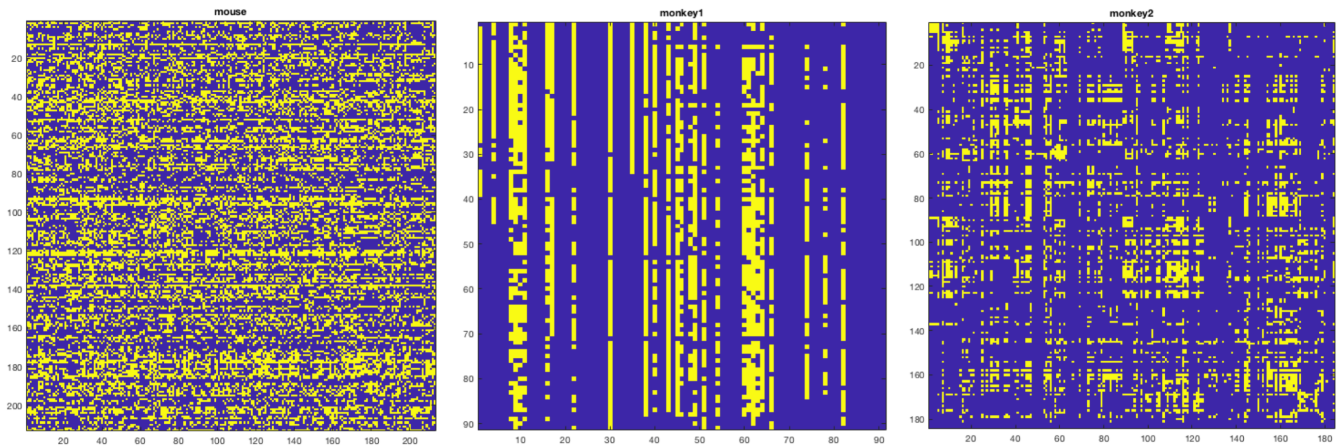

Figure 1. Adjacency matrices of the connectomes tested. *Mouse* includes the mouse whole brain (Oh et al., 2014), while *monkey1* (Markov et al., 2014) and *monkey2* (Bakker et al., 2012) include monkey cortex. All edges (yellow) are directed and have weight 1.

Message-passing Models

In our models, time is discretized and message-passing at each node is synchronous. During a time step, a node can be active or inactive. Each time step t, L new messages, representing message load, are injected into the system at randomly chosen nodes. L is the primary independent variable in this study. Variable loads are to be expected in brain networks, but are most likely restricted to a relatively low range given metabolic constraints. Because the three connectomes have different numbers of nodes N, load is expressed as a percentage of the number of nodes in the network. We also report the case of 1 new message injected per time step in each case. We restrict simulations to the load regime with a number of messages less than or equal to 30% of N. During each time step, a node looks to all incoming messages, including messages that are already in the system and new messages that are scheduled to be injected; if there is more than one message coming in, all are deleted. See pseudocode for the models in Box 1.

Measures of Dynamic Network Activity

Simulations consist of 500 runs of 1000 time steps each. Data from the first 500 time steps is excluded to allow the analysis of equilibrium dynamics only. First, we measured the fraction of nodes active on a given time step. We measured both attempted activity before accounting for collisions, and actual (net) activity after collisions. We also measured the sparseness of the network over time in 5-time step bins using the Treves-Rolls measure of sparseness (Rolls and Tovee, 1994). In particular, we calculate population sparseness by constructing a histogram of the frequency of firing x_n within a time window per node, then sum over i nodes. The Treves-Rolls measure S divides the square of the sum of the distribution by the sum of the squares of the distribution:

$$S = \frac{(\frac{1}{N}\sum_i x_n)^2}{\frac{1}{N}\sum_i x_n^2}$$

This measure ranges between maximal sparseness at 0 (one unit fully active per window, all other units inactive) and minimal sparseness at 1 (for normally-distributed activity). Thus, lower values of Treves-Rolls sparseness indicate more sparse activity. We note that tests of 10-time step windows for the sparseness measure produced similar results. We also calculated lifetime sparseness as above, summing over time instead of over nodes.

Algorithm 1 Information Spreading

```
procedure MESSAGE FEED
    N ← the number of states in the network
    Adj ← the adjacency matrix of the networks
    n_batch ← how many batches of messages will be fed to the network
    load ← number of messages in each batch
    Msg ← load random sender and receiver node pairs
    Senders ← all sender nodes in Msg
    Activity ← n_batch × N matrix of 0
    Activity(1, Senders) ← 1
    Attempts(1, Senders) ← 1
end procedure
procedure MESSAGE DISTRIBUTION
loop:
    for i ← 2, n_batch do
        Activity(i, :) ← Activity(i − 1, :) × Adj
        Msg ← load random sender and receiver node pairs
        Senders ← all sender nodes in Msg
        NewActivity ← 1 × N vector of 0
        NewActivity(Senders) ← 1
        Activity(i, :) ← Activity(i, :) + NewActivity
        Attempts(i, :) ← Activity(i, :)
        Where ← where entries of Activity(i, :) are ≥ 1
        Activity(i, Where) ← 0
        Attempts(i, Where) ← 1
    end for
end procedure
```

Algorithm 2 RandomWalk

```
procedure MESSAGE FEED
    n_batch ← how many batches of messages will be fed to the network
    load ← number of messages in each batch
    Msg ← load random sender and receiver node pairs
    Activity(1, :) ← all sender nodes in Msg
    Attempts(1, :) ← all sender nodes in Msg
end procedure
procedure MESSAGE DISTRIBUTION
loop:
    for i ← 2, n_batch do
        for j ← 1, length of Activity(i − 1, :) do
            PossibNext ← all nodes reachable from Activity(i − 1, j)
            NextMsg(j) ← a randsample from PossibNext
        end for
        Msg ← load random sender and receiver node pairs
        Activity(i, :) ← [NextMsg, Msg]
        Attempts(i, :) ← all unique entries in Activity(i, :)
        Jams ← all repeated elements in Activity(i, :)
        if Jams ≠ ∅ then
            for k = 1, length of Jams do
                Where ← positions of entries in Activity that equal to Jams(k)
                Activity(i, Where) ← 0
            end for
        end if
    end for
end procedure
```

Box 1. Pseudocode of information spreading (IS) and random walk (RW) models.

Comparison to Randomized Networks

To investigate the dependency of network dynamics on the specific wiring patterns of the mammal connectome, we compared network activity under our models to the same models implemented on randomized networks. Randomized networks have the same distribution of in- and out-degree as the empirical networks. This amounts to shuffling the in- and outgoing edges of nodes in a given network (Maslov and Sneppen, 2002). However, the nature of the connectomes in this study complicates the randomization process. When a network is dense, as in the *mouse* connectome and the 29-node core of *monkey1*, network topology changes little after randomization. This problem is compounded in *monkey1* because the full network is low-density as compared to the other two networks. Therefore, results of randomized network comparison are reported in two cases: the *monkey2* connectome, which has intermediate connectivity; and a thresholded *mouse* connectome. Since edge weights in the *mouse* are known (corresponding to tracer volumes) we thresholded the network to exclude edges with a weight below 0.0136 (a value close to the modal weight, and one implied as meaningful by Oh et al., 2014), then we randomized the resulting network. Thresholding of the *mouse* network reduced degree in a way that brought edge density to a level comparable to that of *monkey2* and it allowed for more meaningful randomization (but interestingly, the global topology changed relatively little after removing more than half of the edges, suggesting self-similarity). Note that each trial in the randomized network simulation corresponds to a different randomized network.

Results

Qualitative differences in network dynamics are apparent in visualizations of node activity over time. In Figure 2, we show attempted and actual (net) activity through a representative simulation of the *mouse* connectome, with nodes shown along the vertical axis and the last 500 time steps of the simulation running from top to bottom (L = 10%). The top row shows attempted message-passing per node according to the color scale shown (i.e., the number of messages colliding). The bottom row shows actual (net) activity (0 or 1). In the left-hand column is shown the IS model, and in the right-hand column is shown the RW model. From inspection, the net activity is more "bursty" and sparse in IS while activity is more uniformly distributed across nodes and time in RW.

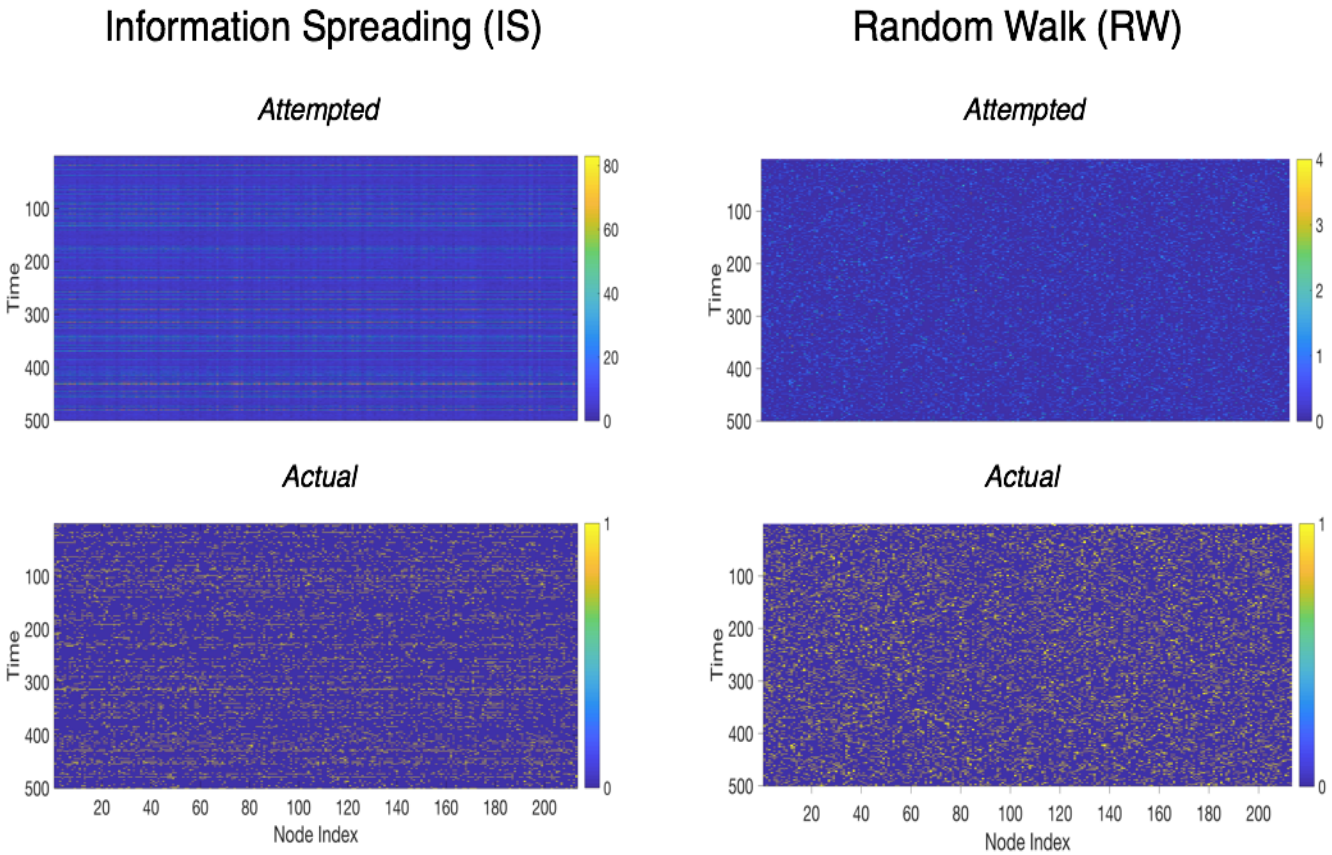

Figure 2. Attempted activity (before collisions) and actual activity (net, following collisions) under IS and RW models in the *mouse1* connectome at 10% load. Horizontal axis corresponds to nodes and vertical axis corresponds to time (running top to bottom). Colors in attempted activity plots represent the number of colliding messages at a given node and time step. Net activity per node per time step is binary.

We can quantify these observations by calculating average net activity and sparseness across simulations. Figure 3 shows average activity (expressed as a fraction of the number of nodes active in a given model, averaged over time and trials) as a function of load L (expressed as a percentage of nodes creating new messages on each time step, except for the left-most tick, which represents injection of one new message per time step). IS models are shown with dashed lines and RW models are shown with solid lines. Though RW models have lower net activity with L = 1 message, activity increases rapidly and monotonically as load increases. In contrast, IS models show relatively uniform activity across loads.

Figure 4 shows sparseness (Treves-Rolls population sparseness) as a function of load. As with activity, population sparseness in a given network is greater (closer to zero) for IS compared to RW at all loads except the lowest load (l message per time step). The same pattern held for lifetime sparseness (See Supplementary Figure 1).

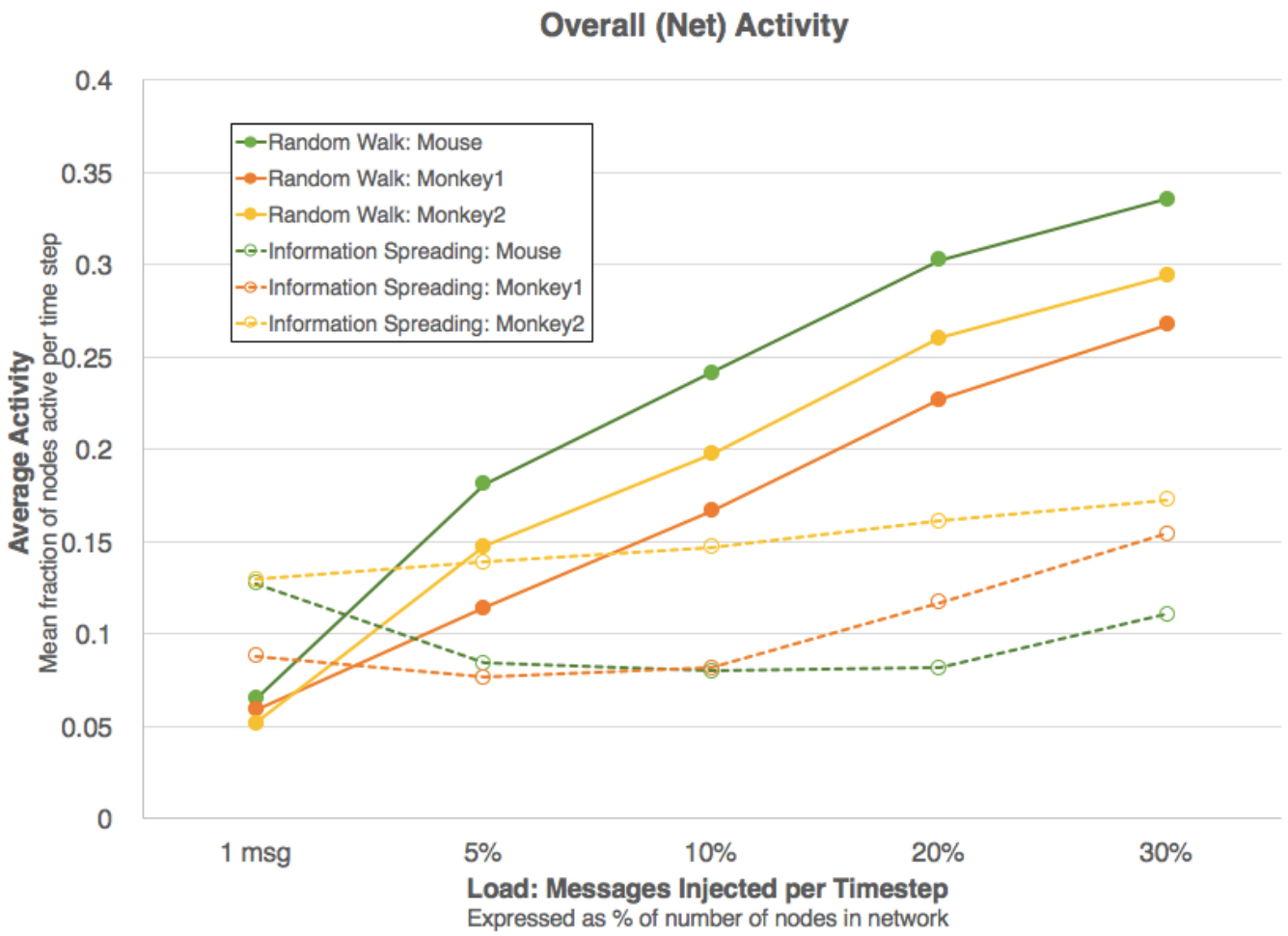

Figure 3. Average (net) activity on each mammal connectome as a function of load under IS (dashed lines, open circles) and RW (solid lines, filled circles) models. In IS, activity remains mostly constant across load and is lower compared to the corresponding value for RW at all load levels except 1 message per time step. Activity in RW increases by more than two octaves across loads tested.

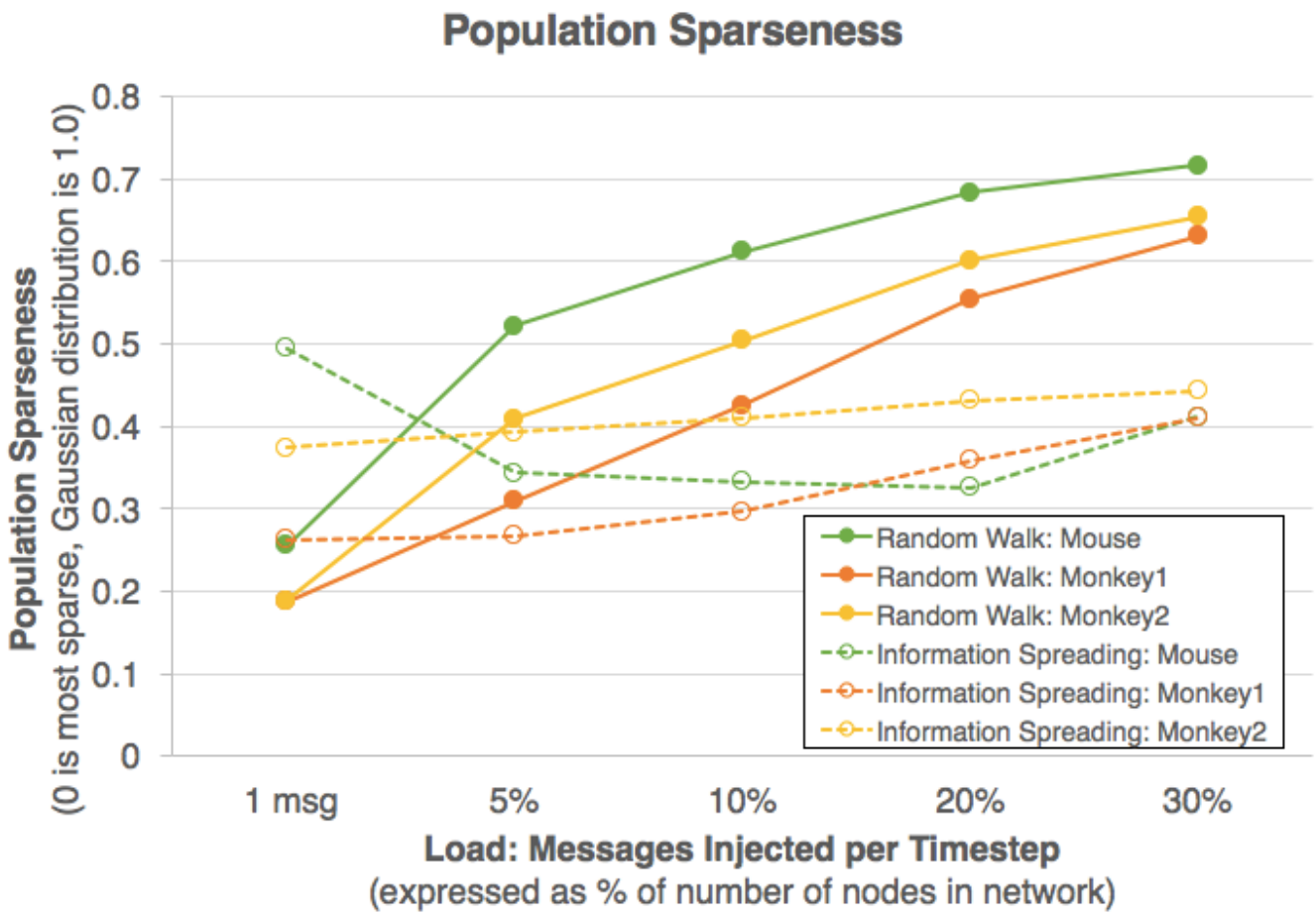

Figure 4. Population sparseness of RW and IS models as a function of load for IS and RW models. As with activity, IS models show relatively constant sparseness across load, with greater sparseness (closer to zero) than RW models at all loads except 1 message per time step. Sparseness in RW decreases by more than a factor of 2.

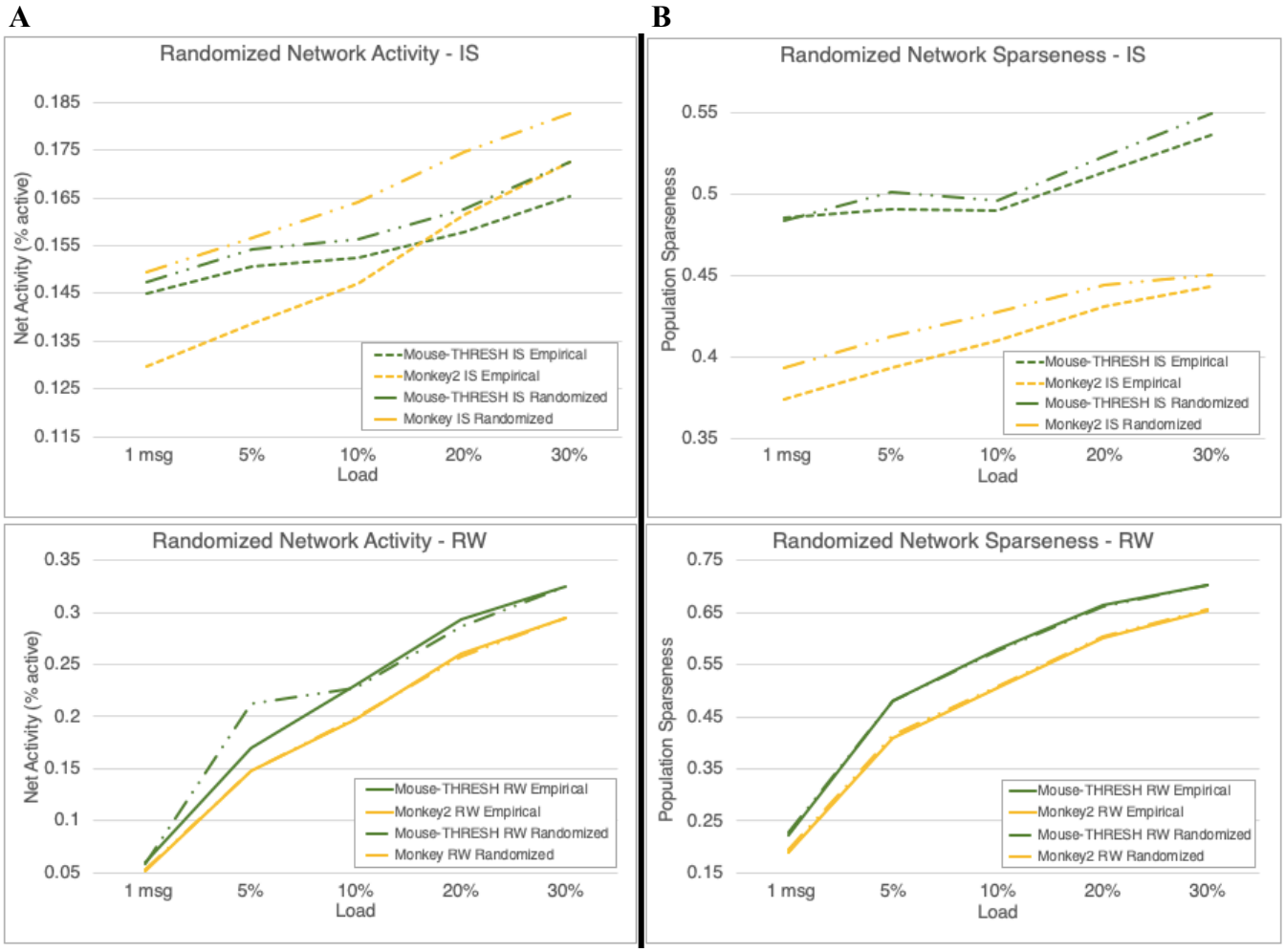

Figure 5. (A) Net activity and (B) population sparseness for empirical and randomized networks of the *monkey2* and the (thresholded) *mouse* under IS and RW models. In IS models, both the empirical *monkey2* and the (thresholded) *mouse* networks show lower activity and greater sparseness (albeit by a small margin) compared to corresponding randomized networks. Randomized RW models show essentially the same behavior as the empirical networks (except in the case of 5% load in the *mouse*, wherein the empirical network is less active).

Results of comparison to randomized networks (Figure 5) in the thresholded *mouse* and in *monkey2* show that IS models have lower activity and greater sparseness in the empirical network at almost all loads tested (albeit by a relatively small margin), whereas RW models show essentially no difference between empirical and randomized networks. This evidence is consistent with the notion that empirical network topology in both species promotes sparse, efficient activity under an information spreading strategy for message routing.

Discussion

We used numerical simulations on the mammal connectome to show that brain networks can exploit the redundancy of short paths by spreading information widely, and in particular by sending multiple copies of an incoming signal. When combined with nonlinear collisions, the network yet achieves globally low

activity and sparseness like what is found in real brains. Importantly, activity and sparseness change relatively little as load increases over a wide range. This result suggests a way that the brain could achieve ongoing sparse activity spread across the entire system, but without deviating too far from homeostatic operation. In contrast, random walk (RW) models with the same collision rule produce substantially higher summed activity. Also, in RW, substantial numbers of nodes are active in a given time window, resulting in less sparseness. Activity and sparseness under RW change substantially as load increases over the same range, with activity increasing more than four-fold, and sparseness decreasing by around a factor of two. Our findings hold in both the monkey cortex and mouse whole brain. Moreover, empirical networks show lower activity and greater sparseness than randomized networks with the same degree distributions—but only in the IS case. Together, these results suggest that nonlinear collision rules and mammal connectome topology could promote sparse, efficient routing of communication traffic across the brain. These rules are biologically plausible, and do not require central control of routing.

Taking a wider perspective, the management and appropriate treatment of signals arising from different sources that impinge on the same target is a critical problem in network neuroscience. Indeed, this is a general problem in cell signaling, and one lacking a general solution. Hancock (2010) writes: "if both hormone X and hormone Y lead to an increase in the concentration of signal Z inside the cell, when signal Z increases, how does the cell know that it is due to the perception of X or Y? It clearly does, but how?" Applying this question to the mesoscale mammal connectome, where all nodes are only a few hops from each other (and where signals are excitatory), the system would seem to need a strategy for managing coincident excitations in order to route signals appropriately.

This conception stands in contrast to standard dynamic models inspired by integrate-and-fire spiking neuron models, wherein the goal of the system is to have many excitatory signals arrive at the same time in order for the unit to be highly activated. Collisions—in the sense of multiple coincident efferents—are traditionally seen as signals to be integrated. This picture mostly holds for single neurons *in vitro* but in large networks in living brains this is less likely to be the case. In part, this is because the mammal brain is sparse. First, only a small fraction of stimuli that could activate neurons are likely to be present at a given time (Vogels and Abbott, 2009). Second, the brain as a whole has severe metabolic constraints that limit simultaneous activity. Thus, it is unlikely that real neural populations have enough energy available for more than a few efferents to be active at a given time. Indeed, it may be desirable for brain networks to generally avoid having more than a small number of inputs to a given node active. By analogy, the Internet also succeeds at managing collisions in part because usage is fundamentally sparse.[1]

Destructive collisions are well established at the level of neurons. Excitatory and inhibitory input signals regularly collide in the brain, causing both to be canceled out. Excitatory signals can also cancel each other when they collide "head on" (this is what is predicted by the Hodgkin-Huxley model; see Scott, 1977). Signals arriving during refractory periods will likewise be destroyed. In addition, Gidon et al. (2020) have recently shown exclusive-OR activity in single *ex vivo* pyramidal neurons in human cortex, a behavior that effectively destroys colliding signals. At the circuit level, gating circuits (e.g., Steriade and Paré, 2007; Gollisch and Meister, 2010), when closed, can also be conceived as destructive collisions between signals. If collisions need to be managed at the single cell level and the level of circuits, the same logic applies at the level of the mesoscale connectome.

[1] The Internet also utilizes arbitration policies such as back-off algorithms (see e.g., Pastor-Satorras and Vespignani, 2011).

Of course, at the mesoscale, inputs arising in different nodes that travel to a target node do not necessarily synapse with the same population within the target node. But collisions cannot be ignored by assuming that signals arriving at a node at the same time do not interact. Such would only be the case if there were dedicated links from sending nodes via a target node to destinations one edge away; this would imply a different network topology (see Supplemental Figure 2). If the brain performs dynamic routing at nodes—and functional demands appear to necessitate this (Olshausen et al., 1993; Graham and Rockmore, 2010)—collisions must be managed.

In separate tests, we found that the structure of mammal brain networks is such that collisions that sum together signals lead to very high activity and very low sparseness in both RW and IS models. This was suggested by tests of a "let one pass" collision rule whereby exactly one message in a collision of k messages is allowed to pass on a randomly chosen outgoing edge. IS models in particular quickly overload with this collision rule (producing data that are not meaningful, and therefore not reported here).

The mesoscale connectome (particularly cortex) has been modeled as aiming to achieve something similar to an integrate-and-fire strategy (i.e., the mechanism at the microscale), such as Kaiser et al. (2007) and Kaiser and Hilgetag (2010). Studies that approach dynamics from this perspective of thresholding also confirm the necessity of keeping global activity at a low and relatively constant level. Kaiser et al. (2010) tested a threshold routing model on surrogate hierarchical networks with biologically plausible parameters, and suggested that brain networks must seek regimes of "limited sustained activity." Using a "spreading" model that is somewhat similar to the IS model, their simulations indicated that specific parameters of activity likelihood promoted limited sustained activity of around 10-20% of nodes. However, many if not most plausible parameter settings led to "dying-out" activity or overload (Kaiser et al. 2007; Kaiser and Hilgetag, 2010). We have shown here that limited sustained activity of around 10-20% of nodes is robustly achieved on mammal connectomes under a nonlinear routing approach across a range of loads based on destructive collisions and information spreading. However, in contrast to Kaiser and Hilgetag (2010), our model shows that this pattern of activity is distributed across the entire network over time, rather than concentrated in a minority of nodes. Moreover, this behavior is achieved robustly on all trials.

The likelihood of collisions in brain networks means that we may need to rethink the range of routing solutions that would be appropriate and effective. Automatic deletion is not necessarily the best solution for managing collisions. The purpose of the present investigation is not to find the optimal solution but rather to ask the question, what would be a good way for the brain to route signals on the mammalian network architecture, given the likelihood of collisions, the need for distributed protocol, and low energy budgets? Our results show that strategies with nonlinear, self-organizing mechanisms that align with basic physiology can generate network activity that offsets collisions, requires no central agent, and operates sparsely in time and over nodes. Much as vision scientists have gained knowledge from asking what a good retinal code would be given the statistical regularities of the environment (e.g., Field, 1994), network neuroscience can advance by asking what a good routing protocol would be in the brain given its architecture and operational demands.

An alternative strategy for managing collisions is to reduce load almost to zero. In this regime, redundancy may not be needed because collisions are rare, and a given path is unlikely to be congested. This is suggested in the finding that RW models achieve lower activity and higher sparseness than IS models at the lowest load (1 message per time step) in all cases. It remains possible that the brain could adopt this kind of extremely sparse strategy (see e.g., Ovsepian, 2015). However, such a system, being so close to floor values, would be restricted to low activity ranges and would be vulnerable to large changes in

dynamics if load varies more than a small amount. In any case, if collisions are indeed very rare, this is also a major problem for models based on integrate-and-fire strategies.

Redundancy in the brain such as correlated firing among neighboring neurons (see e.g., Meytlis, Nichols and Nirenberg, 2007) is typically thought to be something that the system should minimize (Wainwright, 1999; Barlow, 2001). Following Shannon's theory of information, parallel communication channels should seek to reduce mutual information to theoretical minima. But the brain's network topology is not parallel but rather small world-like. In this situation, there is no straightforward way to apply Shannon's theory for arbitrary node-to-node communication (El Gamal and Kim, 2011). Instead, we can look to efficient engineered solutions. On the massive small-world communication network of the Internet, routing that exploits redundancy is fundamental. Individual packets comprising a message can take advantage of these redundant paths. Like the Internet, there is substantial path redundancy in mammal brain networks: brain networks offer many short routes from node to node. These multiple paths could be exploited if neurons or groups of neurons function as routers of information capable of directing particular inputs to particular outputs. In single units, this idea was first termed the *multiplex neuron hypothesis* (Waxman, 1972).

The difficulty of studying nonlinear interactions of brain network communication is due in part to the problem of simulating dynamics explicitly. The challenge lies in tracking every signal's path and interactions. Random walk models and shortest path models admit analytical solutions that capture the likelihood of a set of states of the network. But this is generally possible only when collisions and other nonlinear interactions are ignored. As a result, explicit models capable of accounting for collisions have been relatively little studied to date (see Graham and Hao, 2018). But if explicit models are studied to a greater extent, we will increasingly be able to examine fine-grained activity patterns such as regional differences and correlations in activity.

Whatever solution the mammal brain has adopted for managing collisions, it is likely that it must generate low average activity that is sparsely distributed (and neither can vary too much). Indeed, energy budgets in the brain are so limited that sparse activity may be not only a constraint but also a primary goal of the network's dynamics.

Acknowledgments

We thank Barbara Finlay, David Field, Marc-Thorsten Hütt, and Ao Tang for helpful discussions.

Author Contributions

YH and DG conceptualized the study, designed the framework, performed analyses, interpreted the results, and wrote the manuscript.

Code Availability

Matlab code for performing simulations under IS and RW models will be posted to GitHub.

References

Abdelnour, F., Voss, H. U., and Raj, A. (2014). Network diffusion accurately models the relationship between structural and functional brain connectivity networks. *Neuroimage*, *90*, 335-347.

Aeschbach, D., Matthews, J. R., Postolache, T. T., Jackson, M. A., Giesen, H. A., & Wehr, T. A. (1997). Dynamics of the human EEG during prolonged wakefulness: evidence for frequency-specific circadian and homeostatic influences. *Neuroscience Letters*, *239*(2-3), 121-124.

Attwell, D., and Laughlin, S. B. (2001). An energy budget for signaling in the grey matter of the brain. *Journal of Cerebral Blood Flow and Metabolism*, 21(10), 1133-114.

Avena-Koenigsberger, A., Yan, X., Kolchinsky, A., van den Heuvel, M., Hagmann, P., and Sporns, O. (2019). A spectrum of routing strategies for brain networks. *PLoS Computational Biology*, *15*(3), e1006833.

Bakker, R., Wachtler, T. and Diesmann, M. (2012), CoCoMac 2.0 and the future of tract-tracing databases, *Frontiers in Neuroinformatics*, doi: 10.3389/fninf.2012.00030

Barlow, H. (2001). Redundancy reduction revisited. *Network: Computation in Neural Systems*, *12*(3), 241-253.

Bell, A. J. and Sejnowski, T. J. (1997). The 'independent components' of natural scenes are edge filters. *Vision Research* 37, 3327-3338.

Bullmore, E., and Sporns, O. (2009). Complex brain networks: graph theoretical analysis of structural and functional systems. *Nature Reviews Neuroscience*, *10*(3), 186.

El Gamal, A., & Kim, Y. H. (2011). *Network Information Theory*. Cambridge University Press.

Field, D. J. (1994). What is the goal of sensory coding? *Neural Computation*, *6*(4), 559-601.

Fornito, A., Zalesky, A., and Bullmore, E. (2016). *Fundamentals of Brain Network Analysis*. Academic Press.

Friston, K. J. (2011). Functional and effective connectivity: a review. *Brain Connectivity*, *1*(1), 13-36.

Funahashi, S. (2015). Functions of delay-period activity in the prefrontal cortex and mnemonic scotomas revisited. *Frontiers in Systems Neuroscience*, *9*, 2.

Gidon, A., Zolnik, T. A., Fidzinski, P., Bolduan, F., Papoutsi, A., Poirazi, P., ... & Larkum, M. E. (2020). Dendritic action potentials and computation in human layer 2/3 cortical neurons. *Science*, *367*(6473), 83-87.

Goldman-Rakic, P. S. (1996). Regional and cellular fractionation of working memory. *Proceedings of the National Academy of Sciences*, *93*(24), 13473-13480.

Gollisch, T., & Meister, M. (2010). Eye smarter than scientists believed: neural computations in circuits of the retina. *Neuron*, *65*(2), 150-164.

Graham, D. J. (2014). Routing in the brain. *Frontiers in Computational Neuroscience* 8:44.

Graham, D. J. (2017). Commentary: Building brains that communicate like machines. *Behavioral and Brain Sciences* e266, 37-38.

Graham, D. J., Chandler, D. M. and Field, D. J. (2006). Can the theory of "whitening" explain the center-surround properties of retinal ganglion cell receptive fields? *Vision Research* 46, 2901-2913.

Graham, D. J. and Field, D. J. (2006). Sparse coding in the neocortex. In: *Evolution of Nervous Systems* ed. Jon H. Kaas and Leah A. Krubitzer. Elsevier, Vol. III, pp. 181-187.

Graham, D. J., and Field, D. J. (2009). Natural images: Coding efficiency. *Encyclopedia of Neuroscience*, *6*, 19-27.

Graham, D. J. and Hao, Y. (2018). A selective diffusion model of brain network activity. *Proceedings of the Conference on Cognitive Computational Neuroscience*, 1195. https://ccneuro.org/2018/proceedings/1195.pdf

Graham, D. J. and Rockmore, D. N. (2011). The packet switching brain. *Journal of Cognitive Neuroscience,* 23 (2), 267-276.

Hahn, G., Ponce-Alvarez, A., Deco, G., Aertsen, A., and Kumar, A. (2019). Portraits of communication in neuronal networks. *Nature Reviews Neuroscience*, *20*(2), 117-127.

Hancock, J. T. (2017). *Cell Signalling*. Oxford University Press.

Kaiser, M., Görner, M., and Hilgetag, C. C. (2007). Functional criticality in clustered networks without inhibition. *New J. Phys.* 9, 110.

Kaiser, M., and Hilgetag, C. C. (2010). Optimal hierarchical modular topologies for producing limited sustained activation of neural networks. *Frontiers in Neuroinformatics*, *4*, 8.

Kleinrock, L. (1976) *Queueing Systems, Vol II: Computer Applications*. Vol. 66. New York: Wiley.

Kopell, N. J., Gritton, H. J., Whittington, M. A., and Kramer, M. A. (2014). Beyond the connectome: the dynome. *Neuron*, *83*(6), 1319-1328.

Lennie, P. (2003). The cost of cortical computation. *Current Biology*, 13(6), 493-497.

Levy, W. B., and Baxter, R. A. (1996). Energy efficient neural codes. Neural computation, 8(3), 531-543.

Maslov, S., and Sneppen, K. (2002). Specificity and stability in topology of protein networks. *Science* 296, 910–913

Mišić, B., Sporns, O., and McIntosh, A. R. (2014a). Communication efficiency and congestion of signal traffic in large-scale brain networks. *PLoS Computational Biology* 10(1), e1003427.

Mišić, B., Goñi, J., Betzel, R. F., Sporns, O., and McIntosh, A. R. (2014b). A network convergence zone in the hippocampus. *PLoS Computational Biology* 10(12), e1003982.

Mišić, B., Betzel, R. F., Nematzadeh, A., Goñi, J., Griffa, A., Hagmann, P. and Sporns, O. (2015). Cooperative and competitive spreading dynamics on the human connectome. *Neuron*, 86(6), 1518-1529.

Mišić, V. B., & Mišić, J. (2014). *Machine-to-machine Communications: Architectures, Technology, Standards, and Applications*. CRC Press.

Meytlis, M., Nichols, Z., and Nirenberg, S. (2012). Determining the role of correlated firing in large populations of neurons using white noise and natural scene stimuli. *Vision Research*, *70*, 44-53.

Navlakha, S., Barth, A. L., and Bar-Joseph, Z. (2015). Decreasing-rate pruning optimizes the construction of efficient and robust distributed networks. *PLoS Computational Biology*, 11(7), e1004347.

Navlakha, S., Bar-Joseph, Z., and Barth, A. L. (2017). Network Design and the Brain. *Trends in Cognitive Sciences*.

Olshausen, B. A., Anderson, C. H., and Van Essen, D. C. (1993). A neurobiological model of visual attention and invariant pattern recognition based on dynamic routing of information. *Journal of Neuroscience*, 13(11), 4700-4719.

Olshausen, B. A., and Field, D. J. (1996). Emergence of simple-cell receptive field properties by learning a sparse code for natural images. *Nature*, *381*(6583), 607.

Ovsepian, S. V. (2019). The dark matter of the brain. *Brain Structure and Function*, *224*(3), 973-983.

Pastor-Satorras, R., and Vespignani, A. (2007). *Evolution and Structure of the Internet: A Statistical Physics Approach*. Cambridge University Press.

Poggio, T. (1984). Routing thoughts. Massachusetts Institute of Technology Artificial Intelligence Laboratory Working Paper 258.

Poldrack, R. A. (2015). Is "efficiency" a useful concept in cognitive neuroscience?. *Developmental Cognitive Neuroscience*, *11*, 12-17.

Rolls, E. T. and Tovee, M. J. (1995). Sparseness of the neuronal representation of stimuli in the primate temporal visual cortex. *Journal of Neurophysiology* 73, 713-726.

Rubinov, M., and Sporns, O. (2010). Complex network measures of brain connectivity: uses and interpretations. *Neuroimage*, *52*(3), 1059-1069.

Scott, A. (1977). *Neurophysics*. John Wiley and Sons.

Seguin, C., Razi, A., and Zalesky, A. (2019). Inferring neural signalling directionality from undirected structural connectomes. *Nature Communications*, *10*(1), 1-13.
Sosnovich, A., Grumberg, O., and Nakibly, G. (2017). Formal Black-Box Analysis of Routing Protocol Implementations. *arXiv preprint arXiv:1709.08096*.
Steriade, M., & Paré, D. (2007). *Gating in Cerebral Networks*. Cambridge University Press.
Turrigiano, G. G. (1999). Homeostatic plasticity in neuronal networks: the more things change, the more they stay the same. *Trends in Neurosciences*, *22*(5), 221-227.
Vogels, T. P., and Abbott, L. F. (2009). Gating multiple signals through detailed balance of excitation and inhibition in spiking networks. *Nature Neuroscience*, *12*(4), 483.
Wainwright, M. J. (1999). Visual adaptation as optimal information transmission. *Vision Research*, *39*(23), 3960-3974.
Waxman, S. G. (1972). Regional differentiation of the axon: a review with special reference to the concept of the multiplex neuron. *Brain Research*, *47*(2), 269-288.
Willmore, B., and Tolhurst, D. J. (2001). Characterizing the sparseness of neural codes. *Network: Computation in Neural Systems*, *12*(3), 255-270.

Supplementary Information

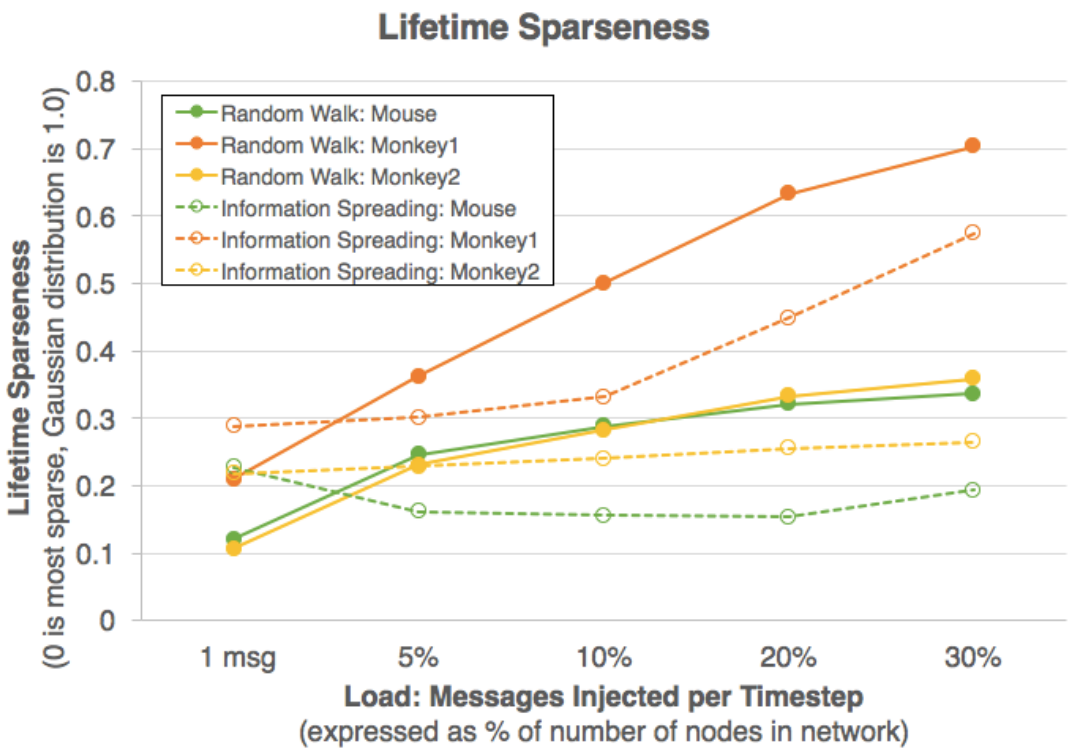

Figure S1. Lifetime sparseness of IS and RW models as a function of load. As with population sparseness, lifetime sparseness is greater (closer to 0) for IS models compared to RW at loads above 1 message per time step.

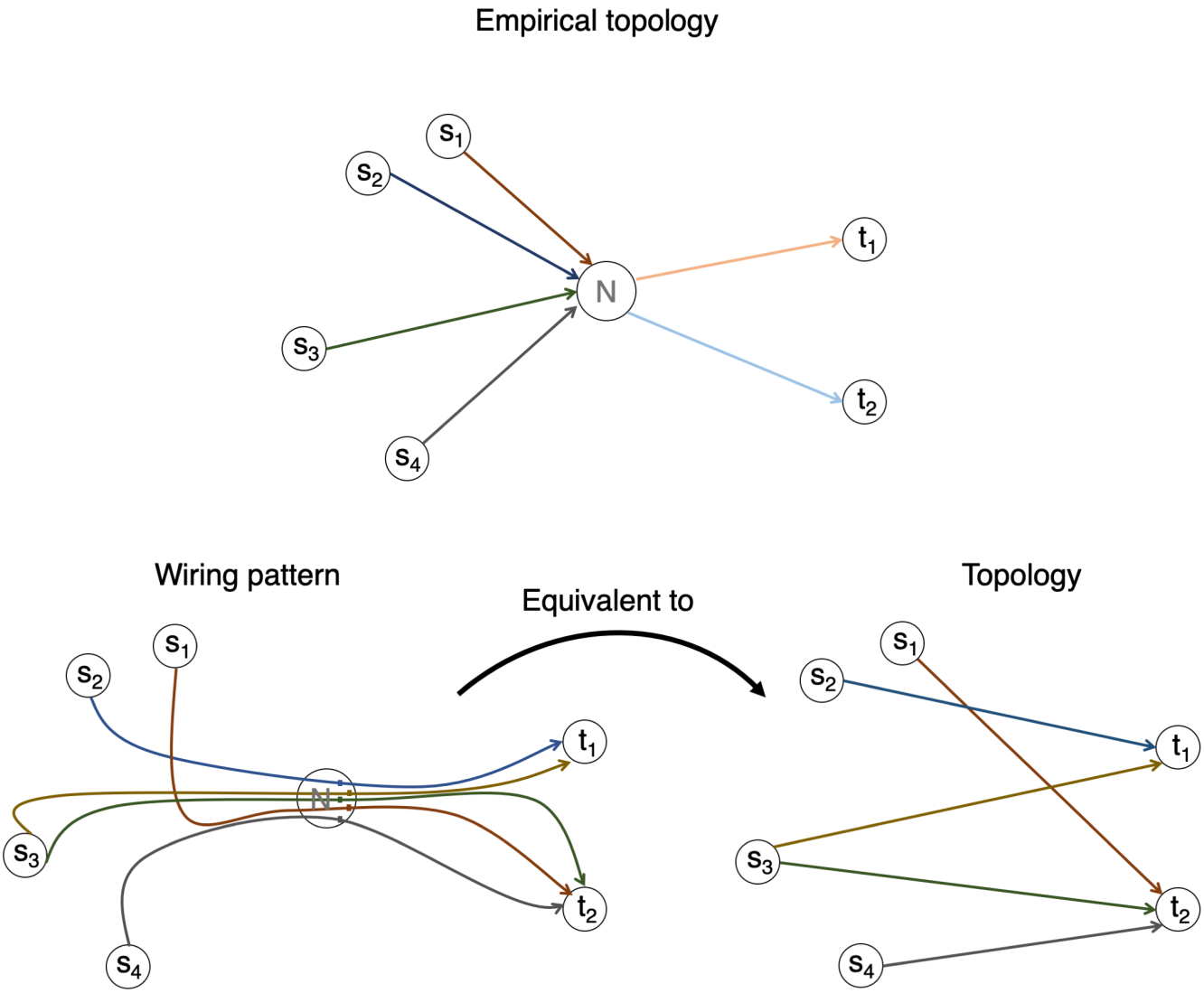

Figure S2. Diagram of brain network topology in comparison to hypothetical wiring pattern within a node (brain region). Top: Nodes s_i have directed edges to node N, and N and has directed edges to t_i. If in reality s_i are wired directly to t_i without interaction within N (bottom left), then topologically N is not a node (vertex). Instead, the network topology is as shown (bottom right). In terms of neurophysiology, inputs to a node interact to a greater or lesser extent due to lateral connections, gap junction interactions, interneurons, convergence, and other mechanisms.