



Human information processing in complex networks

Christopher W. Lynn¹, Lia Papadopoulos¹, Ari E. Kahn^{1,2,3} and Danielle S. Bassett^{1,3,4,5,6,7}

Humans communicate using systems of interconnected stimuli or concepts—from language and music to literature and science—yet it remains unclear how, if at all, the structure of these networks supports the communication of information. Although information theory provides tools to quantify the information produced by a system, traditional metrics do not account for the inefficient ways that humans process this information. Here, we develop an analytical framework to study the information generated by a system as perceived by a human observer. We demonstrate experimentally that this perceived information depends critically on a system’s network topology. Applying our framework to several real networks, we find that they communicate a large amount of information (having high entropy) and do so efficiently (maintaining low divergence from human expectations). Moreover, we show that such efficient communication arises in networks that are simultaneously heterogeneous, with high-degree hubs, and clustered, with tightly connected modules—the two defining features of hierarchical organization. Together, these results suggest that many communication networks are constrained by the pressures of information transmission, and that these pressures select for specific structural features.

Humans receive information in discrete chunks, which transition from one to another—as words in a sentence or notes in a musical progression—to create coherent messages. The networks formed by these chunks (nodes) and transitions (edges) encode the structure of allowed messages, fundamentally governing the ways that we communicate with one another. Although attempts to study the information properties of such transition networks date to the foundation of information theory itself¹, with applications to linguistics^{2,3}, music theory⁴, social and information networks^{5,6}, the Internet⁷ and transportation⁸, fundamental questions concerning the impact of network structure on how humans process information remain unanswered.

The primary difficulty in quantifying the information content of a message is accounting for the human perspective: formally, a message’s information content is not inherent, but rather depends crucially on the receiver’s expectations (or estimated probabilities) of different symbols and stimuli^{1,3,9}. Whereas for computers the probabilities of different symbols are often prescribed, human expectations are biased¹⁰ and differ from person to person³, with measurable consequences for behaviour¹¹ and cognition¹². However, recent advances in psychology and neuroscience have shed light on how humans learn and internally estimate the structure of complex probabilistic systems^{13–18}. Given this progress, it is now possible and compelling to build a framework to quantify human information processing and to consider what types of networks support efficient communication.

Humans perceive information beyond entropy

We set out to study the amount of information a human perceives when observing a sequence of stimuli. Naturally, one might naively expect a human to perceive roughly the same amount of information

as is being produced by a sequence, or its Shannon entropy^{1,9}. Here, to motivate our analytic results, we carry out a set of experiments showing that these two quantities—the information perceived by a human and the information produced by a sequence—differ systematically. To experimentally measure perceived information, we employ a paradigm recently developed in statistical learning^{15–18}, presenting participants with sequences of stimuli on a screen (Fig. 1a) and asking them to respond to each stimulus by pressing the indicated keys on a keyboard (Fig. 1b). Although many real communication systems have long-range correlations, the production of information is traditionally modelled as a Markov process^{1,9}, or equivalently, a random walk on a (possibly weighted, directed) network⁵. Therefore, we assign each stimulus to a node in an underlying network, and we stipulate the order of stimuli within a sequence using random walks (Fig. 1b; Methods). By measuring participants’ reaction times and error rates, we can infer how much information they perceive: slow reactions or many errors reflect surprising transitions (with high perceived information), while fast reactions or few errors indicate expected transitions (with low perceived information)^{11,16,17}.

In a random walk, the probability of transitioning from node (or stimulus) i to a neighbouring node j is $P_{ij} = 1/k_i$, where k_i is the degree of node i . Thus, the amount of information produced by a single transition $i \rightarrow j$ (often referred to as surprisal¹) is given by $-\log P_{ij} = \log k_i$ (Fig. 1d)⁹. Indeed, participants’ behaviour is remarkably well predicted by the information surprisal, with each additional bit of produced information inducing a linear 32 ms increase in reaction times (Fig. 1e) and a 0.3% increase in the number of errors (Supplementary Section 6). However, even if we present participants with networks of constant degree—forcing each transition to produce an identical amount of information—we still

¹Department of Physics and Astronomy, College of Arts and Sciences, University of Pennsylvania, Philadelphia, PA, USA. ²Department of Neuroscience, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ³Department of Bioengineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ⁴Department of Electrical and Systems Engineering, School of Engineering and Applied Science, University of Pennsylvania, Philadelphia, PA, USA. ⁵Department of Neurology, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁶Department of Psychiatry, Perelman School of Medicine, University of Pennsylvania, Philadelphia, PA, USA. ⁷Santa Fe Institute, Santa Fe, NM, USA.

e-mail: dsb@seas.upenn.edu

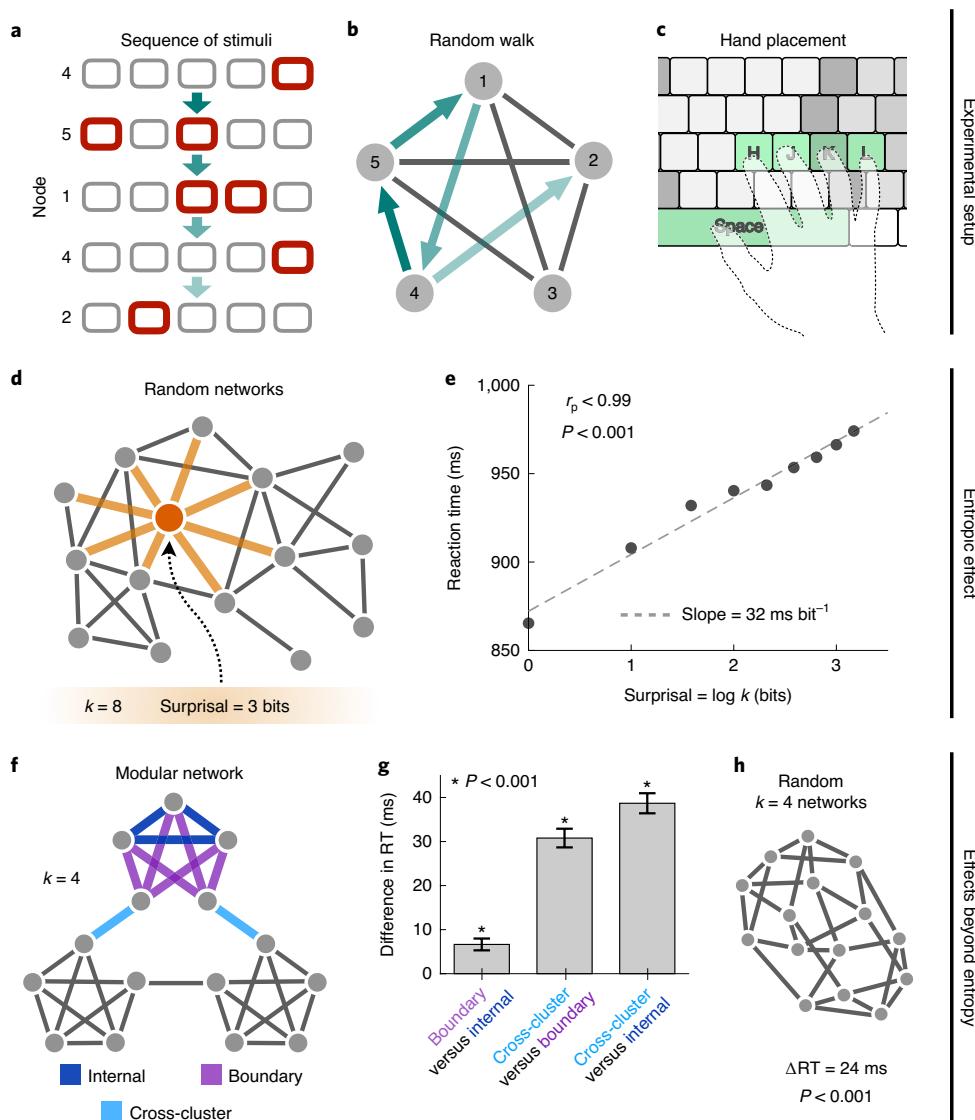


Fig. 1 | Human behavioural experiments reveal the dependence of perceived information on network topology. **a–c**, Experimental setup for our serial reaction time tasks. **a**, Participants are shown sequences of 1,500 stimuli, with each stimulus consisting of five squares with one or two highlighted in red. **b**, The sequential order of stimuli is determined by a random walk on an underlying network. **c**, In response to each stimulus, participants press keys on a keyboard corresponding to the highlighted squares. We use both one- and two-button responses because they allow for networks of size up to $N=15$. To control for the behavioural effects of the different one- and two-button responses, we (1) randomize the assignment of stimuli to nodes for each participant and (2) regress out behavioural dependencies on individual stimuli¹⁶. **d,e**, Effect of produced information on reaction times, referred to as the entropic effect. **d**, For each participant, we draw an Erdős-Renyi random network with $N=15$ nodes and $E=30$ edges; the information produced by a transition $i \rightarrow j$ (or its surprisal) is $\log k_i$, where k_i is the degree of node i . **e**, Reaction times, averaged over all transitions that begin at nodes of a given degree k , are significantly correlated with the produced information $\log k$ (Pearson correlation coefficient $r_p=0.99$, $P<0.001$, $n=177$ participants). **f–h**, Effects of network topology on reaction times after controlling for produced information. **f**, We control for variations in produced information by focusing on networks of constant degree $k=4$, such as the modular network, which contains three distinct types of transitions: those deep within clusters (dark blue), those at the boundaries of clusters (purple) and those between clusters (light blue). **g**, Each type of transition produces reaction times that are distinct from the other two; differences in reaction times and P values are estimated using mixed effects models (two-sided F -test, $n=173$ participants; see Supplementary Section 5). **h**, The difference in reaction times ΔRT between random degree-four networks and the modular network; the modular network yields consistently faster reactions (two-sided F -test, $n=84$ participants). In addition to the population-level results in **e**, **g** and **h**, we also find significant individual variation in participants' sensitivity to network topology (Supplementary Figs. 4 and 5).

discover consistent variations in behaviour that are driven by network topology.

For example, consider the modular network in Fig. 1f, which by symmetry only contains three types of transitions. Each transition produces reaction times and error rates that are distinct from the other two (Fig. 1g), with transitions between or at the boundaries

of clusters generating longer reaction times and more errors (Supplementary Section 6) than those deep within a cluster. In addition to differences in behaviour at the level of individual transitions, we also find overall variations between different networks. Specifically, when compared with random networks of constant degree (Fig. 1g), the modular network yields significantly faster

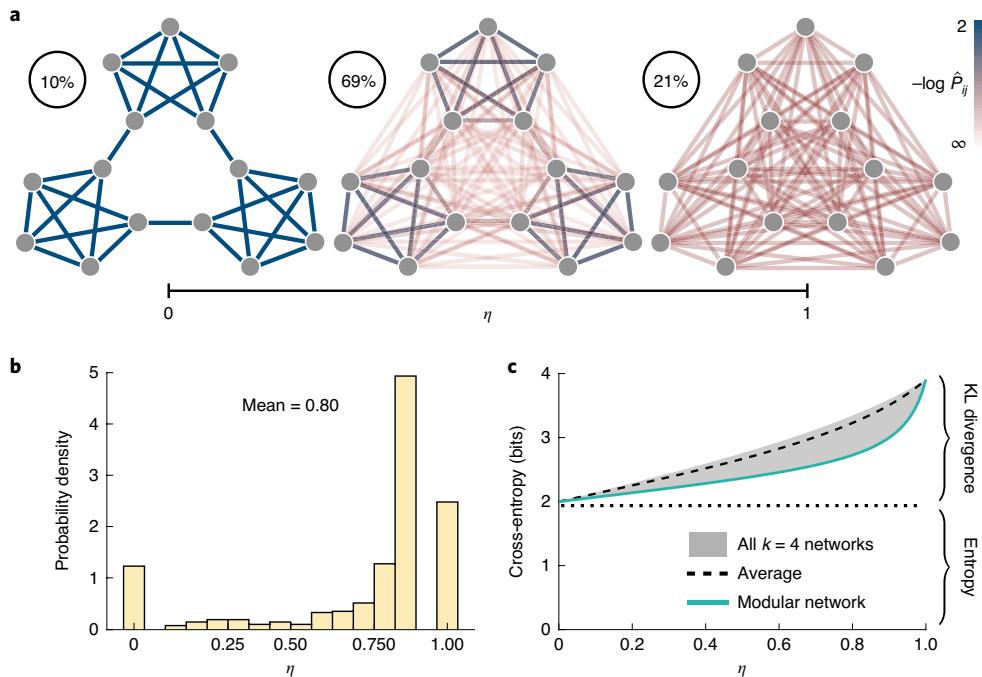


Fig. 2 | Modelling human estimates of transition probabilities. **a**, Illustration of the internal estimates of the transition probabilities \hat{P} in the modular network. For $\eta \rightarrow 0$ (left), the estimates become exact, while for $\eta \rightarrow 1$ (right), the estimates become all-to-all, losing any resemblance to the true network. For intermediate η (centre), transitions within clusters maintain higher probabilities (and therefore lower surprisal) than transitions between clusters, thereby explaining the differences in reaction times in Fig. 1g. Percentages indicate the proportion of participants, across all tasks, belonging to each category. **b**, Distribution of the accuracy parameter η estimated from participants' reaction times (Supplementary Section 4); the distribution is over all 518 completed tasks ($n=434$ participants). **c**, Cross-entropy $S(P, \hat{P})$ as a function of η for all $k=4$ networks of size $N=15$ (shaded region). The modular network (solid line) maintains a lower cross-entropy than the average across all $k=4$ networks (dashed line), thereby explaining the difference in reaction times in Fig. 1h.

reactions (and swifter learning rates; Supplementary Section 7), indicating a decrease in the average perceived information. At the individual level, we find that these network effects vary from person to person (Extended Data Fig. 1) and are often significantly correlated across subjects (Extended Data Fig. 2). Moreover, similar effects have recently been demonstrated across a range of experimental settings¹⁸, including networks of varying size and topology^{16,17,19}; networks with weighted edges^{13,14,20}; time-varying networks^{17,20}; different types of stimuli^{13–15,19–21}; and various behavioural and cognitive measures^{13–15}. Together, these results reveal that humans perceive information—beyond the information produced by a sequence—in a manner that depends systematically on network topology.

Quantifying perceived information using cross-entropy

The differences between perceived information and produced information can be understood as stemming from the inaccuracy of human expectations. As discussed above, given a transition probability matrix P , a transition $i \rightarrow j$ produces $-\log P_{ij}$ bits of information. By contrast, to a person with estimated transition probabilities \hat{P} , the same transition will convey $-\log \hat{P}_{ij}$ bits of information.

Although several models have been proposed for how humans estimate transition probabilities^{14,15,18,20}, converging evidence indicates that humans integrate transitions over time^{17,22–25}. Such temporal integration yields expectations that include higher powers of the transition matrix: $\hat{P} = C \sum_{t=0}^{\infty} f(t) P^{t+1}$, where t is time, $f(t) \geq 0$ is a decreasing function and $C = (\sum_t f(t))^{-1}$ is a normalization constant (we note that \hat{P} is guaranteed to converge if $\sum_t f(t)$ converges). For example, if $f(t) = 1/t!$, then the transition probability estimates \hat{P} are nearly identical to the network communicability^{25–27} (Supplementary Section 4). Here, we focus on the specific choice $f(t) = \eta^t$, where $\eta \in (0,$

1)

represents the inaccuracy of a person's expectations (Fig. 2a). This model can be derived from a number of different cognitive theories—including the temporal context model of episodic memory²², temporal difference learning and the successor representation in reinforcement learning^{23,24}, and the free-energy principle from information theory¹⁷. Inferring η from each participant's reaction times (Fig. 2b; Methods), we find that 10% of participants hold exact estimates of the transition structure ($\eta \rightarrow 0$; Fig. 2a, left), while 21% have expectations that are completely disordered ($\eta \rightarrow 1$; Fig. 2a, right). Importantly, most participants have expectations that lie between these two extremes (Fig. 2a, centre), yielding a decrease in the expected probability of between- versus within-cluster transitions in the modular network. This decrease in expected probability, in turn, gives rise to an increase in perceived information, thereby explaining the observed variations in participants' reaction times and error rates for different parts of the modular network (Fig. 1g).

We are now prepared to study the average perceived information of an entire communication network. Averaging the perceived information of individual transitions over the random walk process, we have $\langle -\log \hat{P}_{ij} \rangle_P = -\sum_{ij} \pi_i P_{ij} \log \hat{P}_{ij}$, where π is the stationary distribution of P . Interestingly, this quantity—known as the cross-entropy $S(P, \hat{P})$ between P and \hat{P} —splits naturally into the entropy $S(P)$, or the average produced information, and the Kullback-Leibler (KL) divergence $D_{KL}(P||\hat{P})$, or the inefficiency of the observer's expectations:

$$\underbrace{\langle -\log \hat{P}_{ij} \rangle_P}_{S(P, \hat{P})} = \underbrace{\langle -\log P_{ij} \rangle_P}_{S(P)} + \underbrace{\langle -\log \frac{\hat{P}_{ij}}{P_{ij}} \rangle_P}_{D_{KL}(P||\hat{P})} \quad (1)$$

Table 1 | Properties of the real communication networks examined in this paper

Type/name	<i>N</i>	<i>E</i>	<i>S</i> ^{real} (bits)	<i>S</i> ^{rand} (bits)	<i>D</i> _{KL} ^{real} (bits)	<i>D</i> _{KL} ^{rand} (bits)
Language (noun transitions)						
Shakespeare	11,234	97,892	6.15	4.16	1.74	2.17
Homer	3,556	23,608	5.25	3.79	1.75	2.12
Plato	2,271	9,796	4.41	3.19	1.74	2.04
Jane Austen	1,994	12,120	4.92	3.66	1.71	2.10
William Blake	370	781	2.59	2.24	1.64	1.77
Miguel de Cervantes	6,090	43,682	5.55	3.89	1.76	2.14
Walt Whitman	4,791	16,526	4.24	2.89	1.76	2.00
Semantic relationships						
Bible	1,707	9,059	4.31	3.48	1.45	2.07
Les Miserables	77	254	3.25	2.82	0.84	1.65
Edinburgh Thesaurus	7,754	226,518	6.26	5.88	2.07	2.21
Roget Thesaurus	904	3,447	3.19	3.02	1.76	1.99
Glossary terms	60	114	2.32	2.09	1.29	1.55
Free On-line Dictionary of Computing	13,274	90,736	4.11	3.83	1.72	2.14
Online Dictionary of Library and Information Science	1,802	12,378	4.59	3.83	1.70	2.11
World Wide Web						
Google internal	12,354	142,296	6.15	4.56	1.35	2.19
Education	2,622	6,065	3.01	2.36	0.92	1.85
Environmental Protection Agencies	2,232	6,876	3.34	2.74	1.75	1.95
Indochina	9,638	45,886	3.88	3.33	0.58	2.08
2004 Election blogs	793	13,484	5.78	5.11	1.36	2.01
Spam	3,796	36,404	5.30	4.30	1.66	2.16
WebBase	6,843	16,374	3.48	2.41	1.09	1.87
Citations						
arXiv Hep-Ph	12,711	139,500	5.02	4.49	1.68	2.19
arXiv Hep-Th	7,464	115,932	5.56	4.98	1.64	2.20
Cora	3,991	16,621	3.50	3.14	1.48	2.04
Digital Bibliography and Library Project	240	858	3.30	2.93	1.37	1.88
Social relationships						
Facebook	13,130	75,562	4.22	3.59	1.78	2.11
arXiv Astr-Ph	17,903	196,972	5.39	4.49	1.41	2.19
Adolescent health	2,155	8,970	3.22	3.14	1.78	2.03
High school	67	267	3.11	3.07	1.15	1.57
Jazz	198	2,742	5.09	4.81	0.94	1.61
Karate club	34	78	2.58	2.32	1.05	1.40
Music (note transitions)						
Thriller—Michael Jackson	67	446	4.03	3.78	0.76	1.38
Hard Day's Night—The Beatles	41	212	3.62	3.42	0.49	1.21
Bohemian Rhapsody—Queen	71	961	5.01	4.77	0.55	0.95
Africa—Toto	39	163	3.41	3.13	0.84	1.29
Sonata No 11—Mozart	55	354	3.91	3.73	0.83	1.28
Sonata No 23—Beethoven	69	900	4.86	4.72	0.65	0.96
Nocturne Op 9-2—Chopin	59	303	3.62	3.42	0.95	1.43
Clavier Fugue 13—Bach	40	143	3.06	2.92	0.89	1.37
Ballade Op 10-1—Brahms	69	670	4.42	4.31	0.80	1.18

For each network, we list its type and name, number of nodes *N* and edges *E*, entropy of the real network *S*^{real} and after randomizing the edges *S*^{rand}, and KL divergence of the real network *D*_{KL}^{real} and after randomization *D*_{KL}^{rand} with η set to the average value 0.80 from our experiments. *S*^{rand} and *D*_{KL}^{rand} are averaged over 100 randomizations. For descriptions of and references for these networks, see Supplementary Table 12.

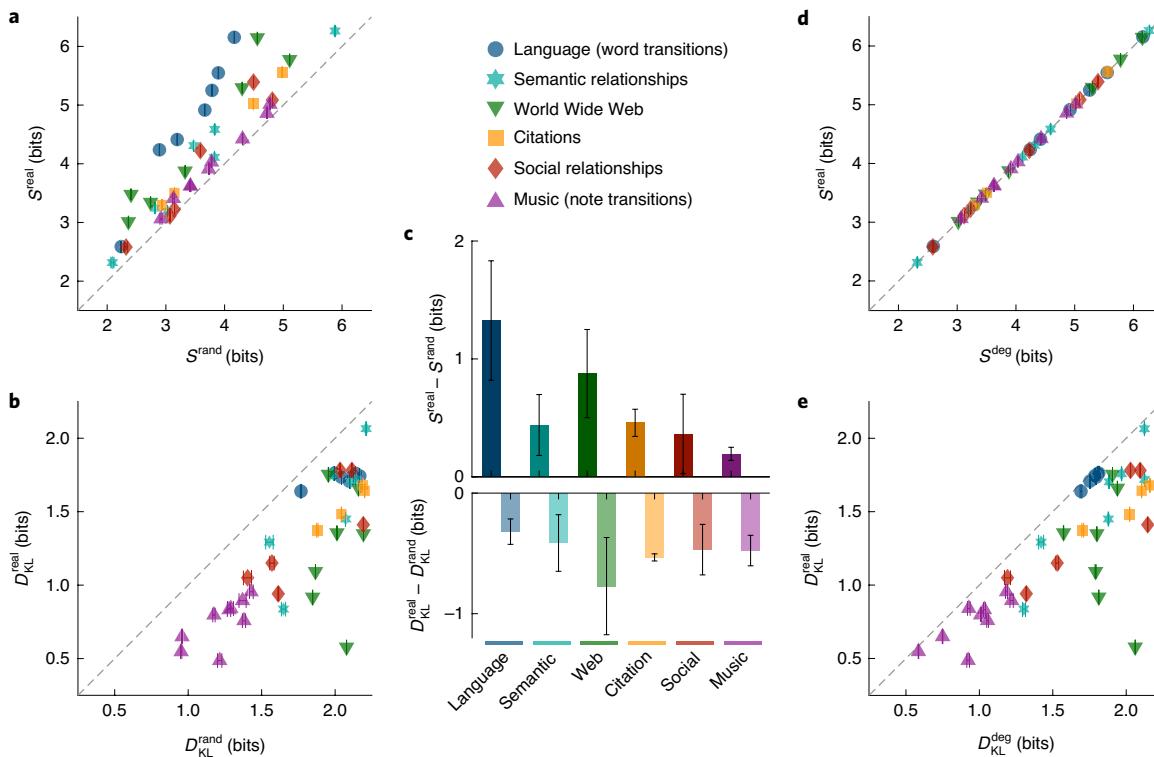


Fig. 3 | The entropy and KL divergence of real communication networks. **a**, Entropy of fully randomized versions of the networks listed in Table 1 (S^{rand}) compared with the true values (S^{real}). **b**, KL divergence of fully randomized versions of the real networks ($D_{\text{KL}}^{\text{rand}}$) compared with the true values ($D_{\text{KL}}^{\text{real}}$). Human expectations \hat{P} are calculated with η set to the average value 0.80 from our experiments; however, the results remain qualitatively the same across all values of η (Supplementary Fig. 6). **c**, Difference between S^{real} and S^{rand} (top) and difference between $D_{\text{KL}}^{\text{real}}$ and $D_{\text{KL}}^{\text{rand}}$ (bottom) for different network types, with error bars indicating standard deviation over networks of each type. **d**, Entropy of degree-preserving randomized networks (S^{deg}) compared with S^{real} . **e**, KL divergence of degree-preserving randomized networks ($D_{\text{KL}}^{\text{deg}}$) compared with $D_{\text{KL}}^{\text{real}}$ with fixed $\eta = 0.80$. In **a**, **b**, **d** and **e**, data points and error bars (standard deviations) are estimated from 100 realizations of the randomized networks, and dashed lines indicate equality between real and randomized quantities. All networks are undirected; for examination of directed versions, see Supplementary Fig. 9.

This relationship has a number of immediate consequences, including the fact that the information a human perceives $S(P, \hat{P})$ is lower-bounded by the information that a system produces $S(P)$ (since $D_{\text{KL}}(P||\hat{P}) \geq 0$). Moreover, inefficiency is minimized when a person's expectations are exact (since $D_{\text{KL}}(P||\hat{P}) = 0$ only when $\hat{P} = P$)⁹. For example, consider the set of degree-four networks from our human experiments (Fig. 1h). While all such networks have identical entropy, their differing topologies induce a range of cross-entropies, which vary as a function of η (Fig. 2c). Notably, the modular graph displays lower cross-entropy than most other degree-four networks (Fig. 2c), thus explaining the observed difference in participants' behaviours (Fig. 1h).

Information properties of real communication networks

Using the framework developed above, we are ultimately interested in characterizing the perceived information generated by real communication systems. The networks chosen (Table 1) have all either evolved or been designed to communicate information through sequences of stimuli (such as words or musical notes) or concepts (such as scientific papers, websites or social interactions). Strikingly, we find that the networks share two consistent properties: they produce large amounts of information (high entropy; Fig. 3a), while maintaining low inefficiency (low KL divergence; Fig. 3b). Specifically, these properties hold relative to completely randomized versions of the networks (Table 1), with η set to the average value 0.8 from our human experiments (Fig. 2b). Interestingly, different network types exhibit these information properties to varying

degrees (Fig. 3c). For example, language networks have the highest entropy but also the highest KL divergence, perhaps reflecting the pressure on language to maximize information rate. Meanwhile, music networks are low in both entropy and KL divergence, possibly mirroring their role as a means for entertainment rather than rapid communication.

If we instead compare the communication networks against randomized versions that preserve node degrees²⁸, we find that the entropy is unchanged (Fig. 3d), indicating that produced information depends only on the degree distribution. By contrast, even compared with these entropy-preserving networks, the KL divergence of real networks remains low (Fig. 3e). We verify that these results largely hold for (1) all values of η (Extended Data Fig. 3), (2) different models of human expectations \hat{P} (Extended Data Figs. 4 and 5), and (3) directed versions of the above networks (Extended Data Fig. 6). Moreover, we find that the information properties of communication networks can vary dramatically in time^{29,30}, with most networks dynamically evolving (for example, over the course of a musical piece or the growth of a social network) to optimize efficient communication—that is, to maximize entropy and minimize divergence from human expectations (Extended Data Figs. 7 and 8).

Finally, to demonstrate that efficient communication is not required by all real communication networks, it is important to consider examples where the results in Fig. 3 break down. We give two such examples in Supplementary Section 11, showing that (1) directed citation networks have markedly low entropy

(Supplementary Fig. 12) and (2) transitions between words of all parts of speech have relatively high KL divergence (Supplementary Fig. 13). However, if we allow transitions to move both forward and backward along citations (as is typical when traversing scientific literature), then citation networks regain their high entropy (Fig. 3a and Extended Data Fig. 9). Similarly, if we focus on ‘content’ words that carry meaning (such as the nouns in Fig. 3) rather than ‘grammatical’ words (such as articles, prepositions and conjunctions)—a common distinction in the study of language networks^{31,32}—then word transitions regain their low KL divergence from human expectations (Fig. 3b,e and Extended Data Fig. 10). Thus, even for networks that appear to have high entropy or low KL divergence, studying the context-specific ways that they transmit information to humans often reveals that efficient communication is maintained.

Efficient communication is driven by hierarchically modular structure

Given the high entropy and low KL divergence displayed by real networks, it is natural to wonder what structural features give rise to these properties. To begin, for undirected networks one can show that $S = \frac{1}{2E} \sum_i k_i \log k_i$, where E is the number of edges, demonstrating that the entropy of a network is determined by its degree sequence (Fig. 3d)³³. It is clear that the entropy grows with increasing node degrees, supporting the intuition that denser networks yield more complex random walks. Moreover, since S is convex in k , the entropy is larger for networks with a small number of high-degree nodes and many low-degree nodes. Interestingly, such heterogeneous structure is observed in human language,³⁴ the Internet,³⁵ social networks³⁶ and scale-free networks³⁵ (although not all networks with heterogeneous degrees are scale free³⁷). To investigate the relationship between a network’s entropy and its degree distribution, we derive a number of analytic results in the thermodynamic limit $N \rightarrow \infty$ (Supplementary Section 12). For example, the entropy of an Erdős-Rényi network is given by $S \approx \log \langle k \rangle$ for large average degree $\langle k \rangle$. For scale-free networks with degree exponent γ (Fig. 4a), we find that $S = \log \langle k \rangle + \frac{1}{\gamma-2} - \log \frac{\gamma-1}{\gamma-2}$, indicating that $\gamma=2$ is a critical exponent since the entropy diverges as $\gamma \rightarrow 2$. Generating ensembles of Erdős-Rényi and scale-free networks, we numerically verify the logarithmic dependence of S on $\langle k \rangle$ (Fig. 4b). Moreover, we find that S increases for decreasing γ (Fig. 4c), suggesting that the entropy grows with increasing degree heterogeneity, which we also confirm numerically (Fig. 4d). This final result reveals that, after controlling for edge density, the entropy is largest for networks with heavy-tailed degree distributions.

In contrast to the entropy, the KL divergence depends on the expectations of an observer. As these expectations become more accurate (that is, as η decreases), we expect $D_{KL}(P||\hat{P})$ to decrease (as in Fig. 2c). But how does the KL divergence depend on network structure? For an unweighted, undirected network with adjacency matrix G , we can expand in the limit of small η to find that

$D_{KL} \approx -\log(1-\eta) - \frac{\eta}{E \ln 2} \sum_i \frac{1}{k_i} \Delta_i$, where $\Delta_i = (G^3)_{ii}/2$ is the number of triangles involving node i (Supplementary Section 13). Therefore, we see that D_{KL} is smaller for networks with a large number of triangles, explaining, for instance, the low KL divergence of the modular network (Figs. 1h and 2c). Indeed, an abundance of triangles is typically associated with modular structure, a ubiquitous feature of real communication networks, from social and scientific interactions^{5,38} to language³⁹ and the Internet⁴⁰. To investigate the impact of modularity on the KL divergence, we derive analytic expressions for D_{KL} that hold for all values of η in the thermodynamic limit (Supplementary Section 13). The KL divergence of an Erdős-Rényi network is given by $D_{KL} = -\log(1-\eta)$. For stochastic block networks with communities of size N_c and a fraction of within-community edges f (Fig. 4e), we find that $D_{KL} = -\log \left[1 - \eta \left(1 - \frac{\langle k \rangle (1-\eta)^f}{N_c (1-\eta)} \right) \right]$. Generating sets of Erdős-Rényi and stochastic block networks, we confirm the analytic predictions that D_{KL} grows with increasing η (Fig. 4f) and decreases for increasing modularity (Fig. 4g) and clustering (Fig. 4h). Therefore, even after controlling for the inaccuracy η of human expectations, we find that modular organization serves to decrease the inefficiency of information transmission.

To attain both the high entropy and low KL divergence observed in real communication systems, it appears that networks must be simultaneously heterogeneous and modular, the two defining features of hierarchical organization⁴¹. To test this hypothesis, we employ a model that combines the heterogeneous degrees of scale-free networks with the modular structure of stochastic block networks (Fig. 5a; see Supplementary Section 14 for an extended description). By adjusting γ and f , we show that these hierarchically modular networks display both a range of entropies (Fig. 5b) and KL divergences (Fig. 5c). In fact, while scale-free networks do not exhibit the low KL divergence of real communication networks nor do stochastic block networks display their high entropy, we find that hierarchically modular networks can attain both properties (Fig. 5d). Taken together, these results indicate that heterogeneity and modularity—precisely the features commonly observed in real communication systems^{5,34–36,38–41}—are both required to achieve high information production and low inefficiency.

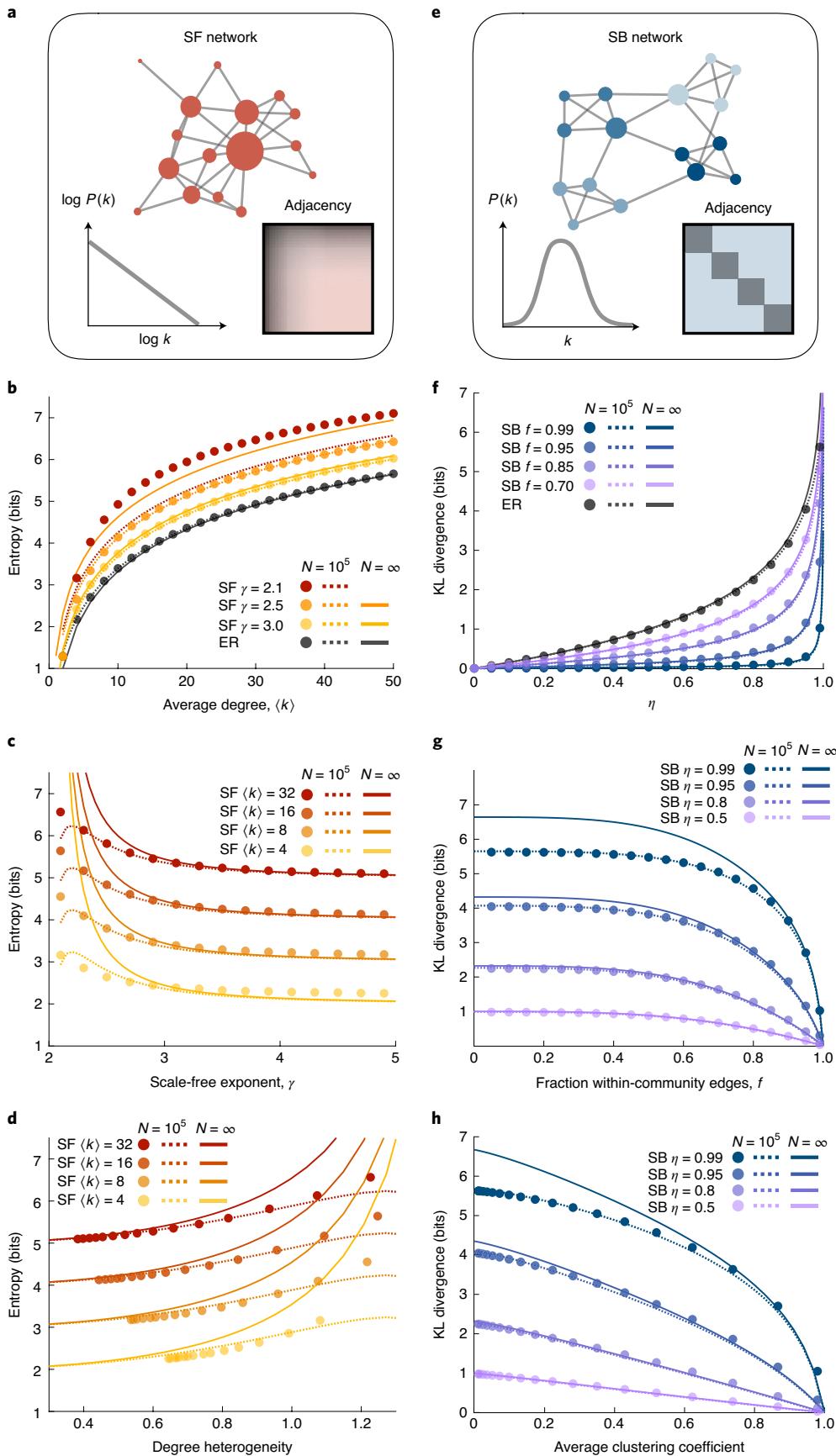
Conclusions and outlook

In this study, we develop tools to quantify the information humans receive from complex networks. We demonstrate experimentally that humans perceive information, beyond the information produced by a sequence, in a way that depends critically on network topology. Moreover, we find that real communication networks support the rapid and efficient transmission of information, and that this efficient communication arises from hierarchical organization. These results raise a number of

Fig. 4 | The impact of network topology on entropy and KL divergence. **a**, Scale-free (SF) network, characterized by a power-law degree distribution and the presence of high-degree hub nodes. **b**, Entropy as a function of the average degree $\langle k \rangle$ for Erdős-Renyi (ER) and SF networks with different scale-free exponents γ . Data points are exact calculations for ER and SF networks generated using the static model⁴⁸ with size $N=10^4$. Lines are derived from the expected degree distributions: dashed lines are numerical results for $N=10^4$ and solid lines are analytic results for $N \rightarrow \infty$ (see Supplementary Section 12 for derivations). Note that the thermodynamic limit for $\gamma=2.1$ does not appear in the displayed range. **c**, Entropy as a function of γ for SF networks with fixed $\langle k \rangle$. In the thermodynamic limit (solid lines), the entropy diverges as $\gamma \rightarrow 2$, and the analytic results are nearly exact for $\gamma > 3$. **d**, Entropy as a function of degree heterogeneity $H = \langle |k_i - k_j| \rangle / \langle k \rangle$, where $\langle |k_i - k_j| \rangle$ is the absolute difference in degrees averaged over all pairs of nodes⁴⁹, for SF networks with fixed $\langle k \rangle$ and variable γ . **e**, Stochastic block (SB) network, characterized by dense connectivity within communities and sparse connectivity between communities. **f**, KL divergence as a function of the accuracy parameter η for ER and SB networks with communities of size $N_c=100$ and different fractions f of within-community edges. Data points are exact calculations for networks with $N=10^4$ and $\langle k \rangle=100$, and lines are analytic calculations for $N=10^4$ (dashed) and $N \rightarrow \infty$ (solid; see Supplementary Section 13 for derivations). **g**, KL divergence as a function of f for SB networks with fixed η . The analytic results are nearly exact for $\eta < 0.8$. **h**, KL divergence as a function of the average clustering coefficient for SB networks with fixed η and variable f .

questions concerning the relationship between human cognition and the structure of communication systems. For example, how have communication networks evolved over time—or

perhaps even co-evolved with the brain⁴²—to facilitate information transmission? Furthermore, how can we design communication systems, from human–technology interfaces⁴³ to



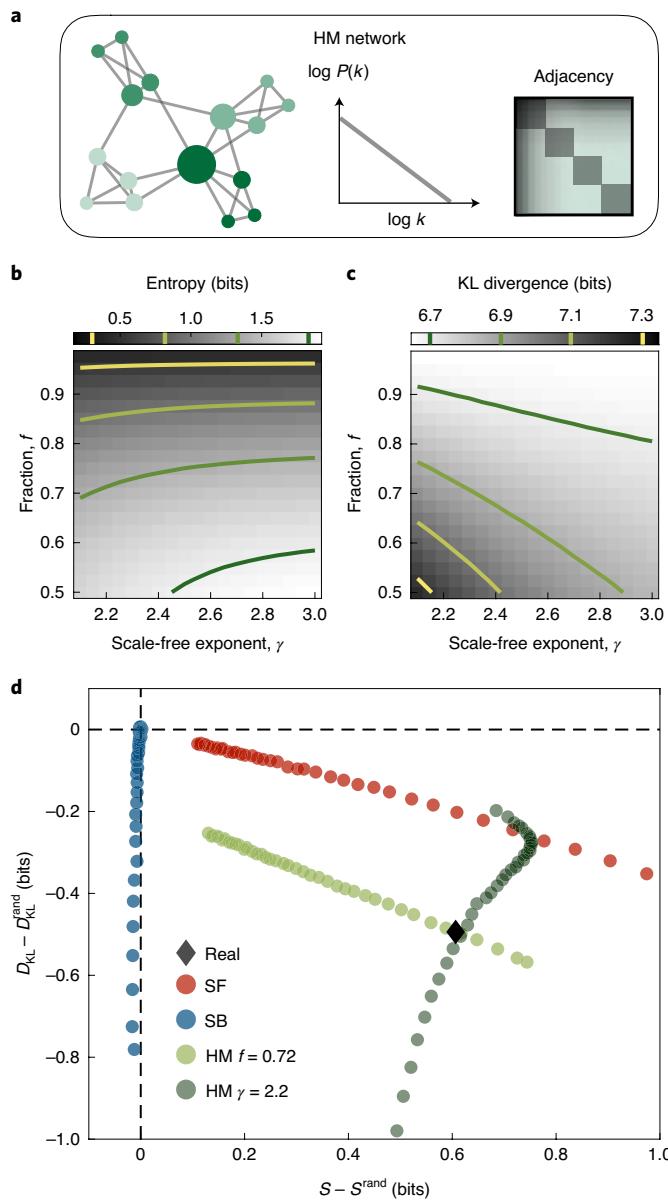


Fig. 5 | Hierarchically modular networks support the efficient communication of information. **a**, Hierarchically modular (HM) network, characterized by a power-law degree distribution and modular structure (Supplementary Section 14). **b**, Entropy as a function of the scale-free exponent γ and the fraction of within-community edges f for HM networks with size $N=10^4$, average degree $\langle k \rangle = 100$ and community size $N_c = 100$. Solid lines denote networks of equal entropy. **c**, KL divergence as a function of γ and f for HM networks with the same size and density as **b** and η set to the average value 0.80 from our experiments (Fig. 2b). Solid lines denote networks of equal KL divergence. **d**, Average entropies and KL divergences of real and model networks compared with fully randomized versions. Data points are averages over the set of networks in Table 1, where for each real network we generate SF networks with variable γ (red), SB networks with communities of size $n \approx \sqrt{N}$ and variable f (blue), and HM networks with $n \approx \sqrt{N}$ and variable γ (fixed $f=0.72$; light green) or variable f (fixed $\gamma=2.2$; dark green), all with N and E equal to the real network. HM networks with $\gamma=2.2$ and $f=0.72$ yield the same average entropy and KL divergence as real communication networks.

classroom lectures⁴⁴, to optimize efficient communication? The framework presented here provides the mathematical tools to begin answering these questions.

To conclude, we highlight a number of ways that our work can be systematically generalized to analyse more realistic communication systems. First, while we model the production of information as a Markov process (equivalently, a random walk), future work should incorporate the long-range dependencies present in many real communication systems^{45,46}. The primary difficulty, however, lies in understanding how humans estimate non-Markov transition structures, with most existing work in statistical learning and artificial grammars focusing on Markov processes^{13–19,21,38,47}. Second, while we have used tools from information theory to quantify the perceived information of a network^{1,9}, these methods do not incorporate the semantic information carried by individual nodes (for example, words, notes, concepts)^{2,3}. Thus, to improve our understanding of real-world communication systems, future progress will require important interdisciplinary efforts from both cognitive scientists (to study how humans estimate non-Markov structures) and information theorists (to quantify semantic information in human contexts).

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at <https://doi.org/10.1038/s41567-020-0924-7>.

Received: 17 July 2019; Accepted: 28 April 2020;

Published online: 15 June 2020

References

- Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
- Bar-Hillel, Y. & Carnap, R. Semantic information. *Br. J. Phil. Sci.* **4**, 147–157 (1953).
- Dretske, F. I. *Knowledge and the Flow of Information* (MIT Press, 1981).
- Cohen, J. E. Information theory and music. *Behav. Sci.* **7**, 137–163 (1962).
- Rosvall, M. & Bergstrom, C. T. Maps of random walks on complex networks reveal community structure. *Proc. Natl Acad. Sci. USA* **105**, 1118–1123 (2008).
- Gómez-Gardeñes, J. & Latora, V. Entropy rate of diffusion processes on complex networks. *Phys. Rev. E* **78**, 065102 (2008).
- Liben-Nowell, D. & Kleinberg, J. Tracing information flow on a global scale using Internet chain-letter data. *Proc. Natl Acad. Sci. USA* **105**, 4633–4638 (2008).
- Rosvall, M., Trusina, A., Minnhagen, P. & Sneppen, K. Networks and cities: an information perspective. *Phys. Rev. Lett.* **94**, 028701 (2005).
- Cover, T. M. & Thomas, J. A. *Elements of Information Theory* (John Wiley & Sons, 2012).
- Hilbert, M. Toward a synthesis of cognitive biases: how noisy information processing can bias human decision making. *Psychol. Bull.* **138**, 211–237 (2012).
- Laming, D. R. J. *Information Theory of Choice-reaction Times* (Academic Press, 1968).
- Koechlin, E. & Hyafil, A. Anterior prefrontal function and the limits of human decision-making. *Science* **318**, 594–598 (2007).
- Saffran, J. R., Aslin, R. N. & Newport, E. L. Statistical learning by 8-month-old infants. *Science* **274**, 1926–1928 (1996).
- Dehaene, S., Meyniel, F., Wacongne, C., Wang, L. & Pallier, C. The neural representation of sequences: from transition probabilities to algebraic patterns and linguistic trees. *Neuron* **88**, 2–19 (2015).
- Schapiro, A. C., Rogers, T. T., Cordova, N. I., Turk-Browne, N. B. & Botvinick, M. M. Neural representations of events arise from temporal community structure. *Nat. Neurosci.* **16**, 486–492 (2013).
- Kahn, A. E., Karuza, E. A., Vettel, J. M. & Bassett, D. S. Network constraints on learnability of probabilistic motor sequences. *Nat. Hum. Behav.* **2**, 936–947 (2018).
- Lynn, C. W., Kahn, A. E., Nyema, N. & Bassett, D. S. Abstract representations of events arise from mental errors in learning and memory. *Nat. Commun.* **11**, 2313 (2020).
- Lynn, C. W. & Bassett, D. S. How humans learn and represent networks. *Proc. Natl Acad. Sci. USA* (in the press).

19. Karuza, E. A., Kahn, A. E. & Bassett, D. S. Human sensitivity to community structure is robust to topological variation. *Complexity* <https://doi.org/10.1155/2019/8379321> (2019).
20. Meyniel, F., Maheu, M. & Dehaene, S. Human inferences about sequences: a minimal transition probability model. *PLoS Comput. Biol.* **12**, e1005260 (2016).
21. Tompson, S. H., Kahn, A. E., Falk, E. B., Vettel, J. M. & Bassett, D. S. Individual differences in learning social and nonsocial network structures. *J. Exp. Psychol. Learn. Mem. Cogn.* **45**, 253–271 (2019).
22. Howard, M. W. & Kahana, M. J. A distributed representation of temporal context. *J. Math. Psychol.* **46**, 269–299 (2002).
23. Dayan, P. Improving generalization for temporal difference learning: the successor representation. *Neural Comput.* **5**, 613–624 (1993).
24. Gershman, S. J., Moore, C. D., Todd, M. T., Norman, K. A. & Sederberg, P. B. The successor representation and temporal context. *Neural Comput.* **24**, 1553–1568 (2012).
25. Garvert, M. M., Dolan, R. J. & Behrens, T. E. A map of abstract relational knowledge in the human hippocampal-entorhinal cortex. *Elife* **6**, e17086 (2017).
26. Estrada, E. & Hatano, N. Communicability in complex networks. *Phys. Rev. E* **77**, 036111 (2008).
27. Estrada, E., Hatano, N. & Benzi, M. The physics of communicability in complex networks. *Phys. Rep.* **514**, 89–119 (2012).
28. Maslov, S. & Sneppen, K. Specificity and stability in topology of protein networks. *Science* **296**, 910–913 (2002).
29. Derex, M. & Boyd, R. The foundations of the human cultural niche. *Nat. Commun.* **6**, 8398 (2015).
30. Momennejad, I., Duker, A. & Coman, A. Bridge ties bind collective memories. *Nat. Commun.* **10**, 1578 (2019).
31. Milo, R. et al. Superfamilies of evolved and designed networks. *Science* **303**, 1538–1542 (2004).
32. Foster, J. G., Foster, D. V., Grassberger, P. & Paczuski, M. Edge direction and the structure of networks. *Proc. Natl Acad. Sci. USA* **107**, 10815–10820 (2010).
33. Burda, Z., Duda, J., Luck, J.-M. & Waclaw, B. Localization of the maximal entropy random walk. *Phys. Rev. Lett.* **102**, 160602 (2009).
34. Cancho, R. F. I. & Solé, R. V. The small world of human language. *Proc. R. Soc. Lond. B* **268**, 2261–2265 (2001).
35. Barabási, A.-L. & Albert, R. Emergence of scaling in random networks. *Science* **286**, 509–512 (1999).
36. Newman, M. E. The structure of scientific collaboration networks. *Proc. Natl Acad. Sci. USA* **98**, 404–409 (2001).
37. Stumpf, M. P. & Porter, M. A. Critical truths about power laws. *Science* **335**, 665–666 (2012).
38. Girvan, M. & Newman, M. E. Community structure in social and biological networks. *Proc. Natl Acad. Sci. USA* **99**, 7821–7826 (2002).
39. Motter, A. E., De Moura, A. P., Lai, Y.-C. & Dasgupta, P. Topology of the conceptual network of language. *Phys. Rev. E* **65**, 065102 (2002).
40. Eriksen, K. A., Simonsen, I., Maslov, S. & Sneppen, K. Modularity and extreme edges of the Internet. *Phys. Rev. Lett.* **90**, 148701 (2003).
41. Ravasz, E. & Barabási, A.-L. Hierarchical organization in complex networks. *Phys. Rev. E* **67**, 026112 (2003).
42. Deacon, T. W. *The Symbolic Species: The Co-evolution of Language and the Brain* (WW Norton, 1998).
43. Dix, A. *Human-Computer Interaction* (Springer, 2009).
44. Hayes, A. F. *Statistical Methods for Communication Science* (Routledge, 2009).
45. Brown, P. F., Desouza, P. V., Mercer, R. L., Pietra, V. J. D. & Lai, J. C. Class-based n-gram models of natural language. *Comput. Linguist.* **18**, 467–479 (1992).
46. Pachet, F., Roy, P. & Barbieri, G. Finite-length Markov processes with constraints. In *Twenty-Second International Joint Conference on Artificial Intelligence* (ed. Walsh, T.) 635–642 (AAAI, 2011).
47. Meyniel, F. & Dehaene, S. Brain networks for confidence weighting and hierarchical inference during probabilistic learning. *Proc. Natl Acad. Sci. USA* **114**, E3859–E3868 (2017).
48. Goh, K.-I., Kahng, B. & Kim, D. Universal behavior of load distribution in scale-free networks. *Phys. Rev. Lett.* **87**, 278701 (2001).
49. Liu, Y.-Y., Slotine, J.-J. & Barabási, A.-L. Controllability of complex networks. *Nature* **473**, 167–173 (2011).

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

© The Author(s), under exclusive licence to Springer Nature Limited 2020

Methods

Experimental setup. Participants performed a self-paced serial reaction time task using a computer screen and keyboard. Each stimulus was presented as a horizontal row of five grey squares; all five squares were shown at all times. The squares corresponded spatially with the keys ‘Space’, ‘H’, ‘J’, ‘K’ and ‘L’ (Fig. 1c). To indicate a target key or pair of keys for the participant to press, the corresponding squares would become outlined in red (Fig. 1a). When participants pressed the correct key combination, the squares on the screen would immediately display the next stimulus. If an incorrect key or pair of keys was pressed, the message ‘Error!’ was displayed on the screen below the stimulus and remained until the participant pressed the correct key(s). The order in which stimuli were presented to each participant was determined by a random walk on a network of $N=15$ nodes. For each participant, one of the 15 key combinations was randomly assigned to each node in the network.

In the first experiment, each participant was assigned an Erdős–Rényi network with $E=30$ edges. In the second experiment, all participants responded to sequences of stimuli drawn from the modular network (Fig. 1f), which has the same number of nodes and edges. We remark that each node in the modular network is connected to four other nodes, so the entropy of each transition was a constant $-\log \frac{1}{4} = 2$ bits. Some participants performed both of the first two experiments in back-to-back stages, with the order of the experiments counterbalanced across participants. In the third experiment, participants underwent two stages. In one stage, participants responded to stimuli drawn from the modular network, while in the other stage, each participant was assigned a random $k=4$ network. The order of the two stages was counterbalanced. For each stage of each experiment, participants responded to sequences of 1,500 stimuli.

Experimental procedures. All participants provided informed consent in writing and experimental methods were approved by the Institutional Review Board of the University of Pennsylvania. In total, we recruited 363 unique participants to complete our studies on Amazon’s Mechanical Turk: 106 completed just the first experiment, 102 completed just the second experiment, 71 completed both the first and second experiments in back-to-back stages, and 84 completed the third experiment. Worker IDs were used to exclude duplicate participants between experiments, and all participants were financially remunerated for their time. In the first two experiments, participants were paid US\$3–11 for up to an estimated 30–60 min: US\$3 per network for up to two networks, US\$2 per network for correctly responding on at least 90% of the trials and US\$1 for completing two stages. In the third experiment, participants were paid up to US\$9 for an estimated 60 min: US\$5 for completing the experiment and US\$2 for correctly responding on at least 90% of the trials on each stage.

Data analysis. To make inferences about participants’ internal expectations based on their reaction times, we excluded all trials in which participants responded incorrectly. We also excluded reaction times that were implausible, either three standard deviations from a participant’s mean reaction time, below 100 ms, or over 3,500 ms.

Measuring the effects of topology on reaction times. To estimate the effects of network topology on participants’ reaction times, one must overcome large interparticipant variability. To do so, we used linear mixed effects models, which have become prominent in human research where many measurements are made for each participant⁵⁰. Compared with standard linear models, mixed effects models allow for differentiation between effects that are participant specific and those that are representative of the prototypical individual in our experiments. Here all models were fit using the `fitlme` function in MATLAB (R2018a), and random effects were chosen as the maximal structure that (1) allowed the model to converge and (2) did not include effects whose 95% confidence intervals overlapped with zero. In what follows, when referring to our mixed effects models, we employ the standard R notation.

For the first experiment, to measure the impact of entropy on reaction times (Fig. 1e), we regressed out a number of biomechanical dependencies: (1) variability due to the different button combinations, (2) the natural quickening of reactions with trial number and (3) the change in reaction times between stages. We also regressed out the effects of recency on participants’ reaction times. Specifically, we fit a mixed effects model with the formula $RT \approx \log(\text{Trial}) \times \text{Stage} + \text{Target} + \text{Recency} + (1 + \log(\text{Trial})) \times \text{Stage} + \text{Recency} | \text{ID}$, where RT is the reaction time, Trial is the trial number (it is common to consider $\log(\text{Trial})$ rather than the trial number itself^{6,17}), Stage is the stage of the experiment, Target is the target button combination, Recency is the number of trials since the last instance of the current stimulus and ID is each participant’s unique ID.

For the second experiment, to measure differences in reaction times between transitions in the modular network (Fig. 1g), we fit a mixed effects model of the form $RT \approx \log(\text{Trial}) \times \text{Stage} + \text{Target} + \text{Recency} + \text{Trans_Type} + (1 + \log(\text{Trial})) \times \text{Stage} + \text{Recency} | \text{ID}$, where Trans_Type is a dummy variable representing the type of transition (Fig. 1g) and the other variables are defined above. The three models for the three different comparisons are summarized in Supplementary Tables 2–4.

For the third experiment, to measure the difference in reaction times between the modular network and random $k=4$ networks (Fig. 1h), we fit a mixed effects

model of the form $RT \approx \log(\text{Trial}) \times \text{Stage} + \text{Target} + \text{Recency} + \text{Graph} + (1 + \log(\text{Trial})) \times \text{Stage} + \text{Recency} | \text{ID}$, where Graph is a dummy variable representing the type of network (either modular or random $k=4$). This model is summarized in Supplementary Table 5.

Estimating η values. Given a choice for the parameter η , and given a sequence of past nodes x_1, \dots, x_{t-1} , the internal expectation of the next node x_t is predicted to be \hat{P}_{x_{t-1}, x_t} . We predict participants’ reaction times $r(t)$ using the linear model $\hat{r}(t) = r_0 - r_1 \log \hat{P}_{x_{t-1}, x_t}$, where $-\log \hat{P}_{x_{t-1}, x_t}$ is the predicted perceived information at time t . Before estimating η , r_0 and r_1 , we regress out participants’ biomechanical dependencies using the mixed effects model $RT \approx \log(\text{Trial}) \times \text{Stage} + \text{Target} + \text{Recency} + (1 + \log(\text{Trial})) \times \text{Stage} + \text{Recency} | \text{ID}$, where all variables are defined above. Then, to estimate the model parameters that best describe a participant’s reactions, we minimize the root-mean-square error (RMSE) with respect to each participant’s reaction times. We note that, given a choice for η , the linear parameters r_0 and r_1 can be calculated analytically. Thus, the estimation problem can be restated as a one-dimensional minimization problem; that is, minimizing RMSE with respect to η . To find the global minimum, we began by calculating RMSE along 101 values for η between 0 and 1. Then, starting at the minimum value of this search, we performed gradient descent until the gradient $\frac{\partial \text{RMSE}}{\partial \eta}$ fell below an absolute value of 10^{-6} . The resulting distribution for η over participants are shown in Fig. 2b. For more details, see Supplementary Section 4.

Reporting Summary. Further information on research design is available in the Nature Research Reporting Summary linked to this article.

Data availability

Source data for Fig. 1, Supplementary Figs. 3–5 and Supplementary Tables 1–11 are provided in Supplementary Data File 1. Source data for Fig. 2 and Supplementary Fig. 1 are provided in Supplementary Data File 2. Source data for the networks in Fig. 3, Table 1 and Supplementary Figs. 6–9 are either publicly available or provided in Supplementary Data File 3 (see Supplementary Table 12 for details).

Code availability

The code that supports the findings of this study is available from the corresponding author upon reasonable request.

References

50. Schall, R. Estimation in generalized linear models with random effects. *Biometrika* **78**, 719–727 (1991).

Acknowledgements

We thank E. Horsley, H. Ju, D. Lydon-Staley, S. Patankar, P. Srivastava and E. Teich for feedback on earlier versions of the manuscript. We thank D. Zhou for providing the code used to parse the texts. D.S.B., C.W.L. and A.E.K. acknowledge support from the John D. and Catherine T. MacArthur Foundation, the Alfred P. Sloan Foundation, the ISI Foundation, the Paul G. Allen Family Foundation, the Army Research Laboratory (W911NF-10-2-0022), the Army Research Office (Bassett-W911NF-14-1-0679, Grafton-W911NF-16-1-0474, DCIST-W911NF-17-2-0181), the Office of Naval Research, the National Institute of Mental Health (2-R01-DC-009209-11, R01-MH112847, R01-MH107235, R21-MH-106799), the National Institute of Child Health and Human Development (1R01HD086888-01), National Institute of Neurological Disorders and Stroke (R01 NS099348) and the National Science Foundation (BCS-1441502, BCS-1430087, NSF PHY-1554488 and BCS-1631550). L.P. is supported by an NSF Graduate Research Fellowship. The content is solely the responsibility of the authors and does not necessarily represent the official views of any of the funding agencies.

Author contributions

C.W.L. and D.S.B. conceived the project. C.W.L. designed the framework and performed the analysis. C.W.L. and A.E.K. performed the human experiments. C.W.L. wrote the manuscript and Supplementary Information. L.P., A.E.K. and D.S.B. edited the manuscript and Supplementary Information.

Competing interests

The authors declare no competing interests.

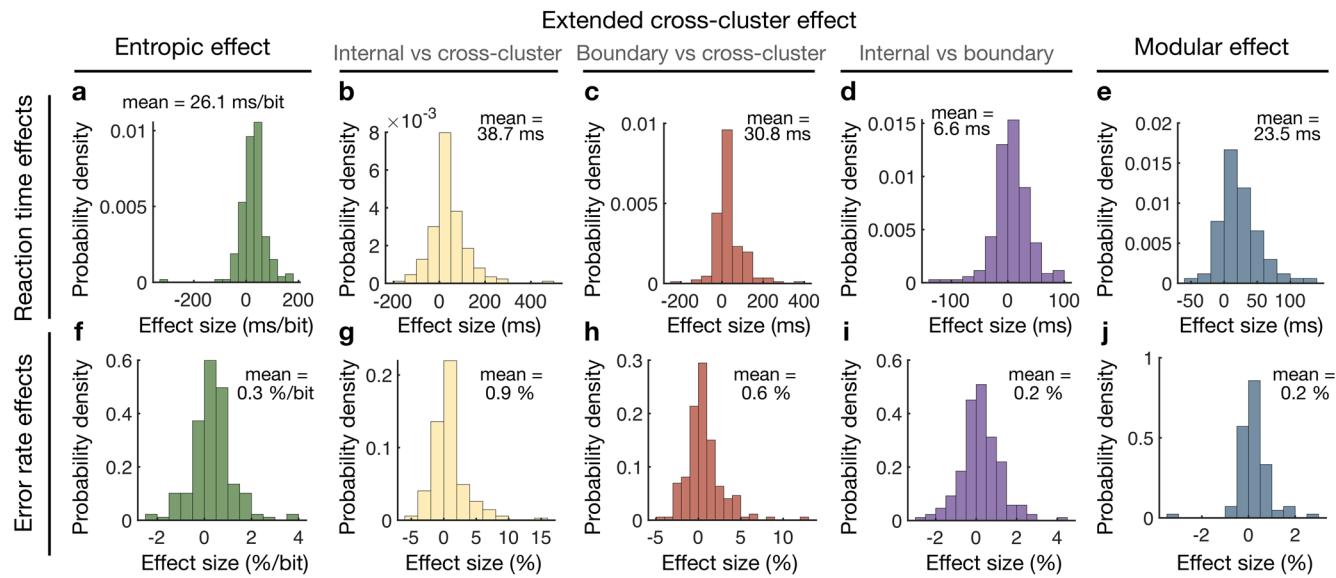
Additional information

Extended data is available for this paper at <https://doi.org/10.1038/s41567-020-0924-7>.

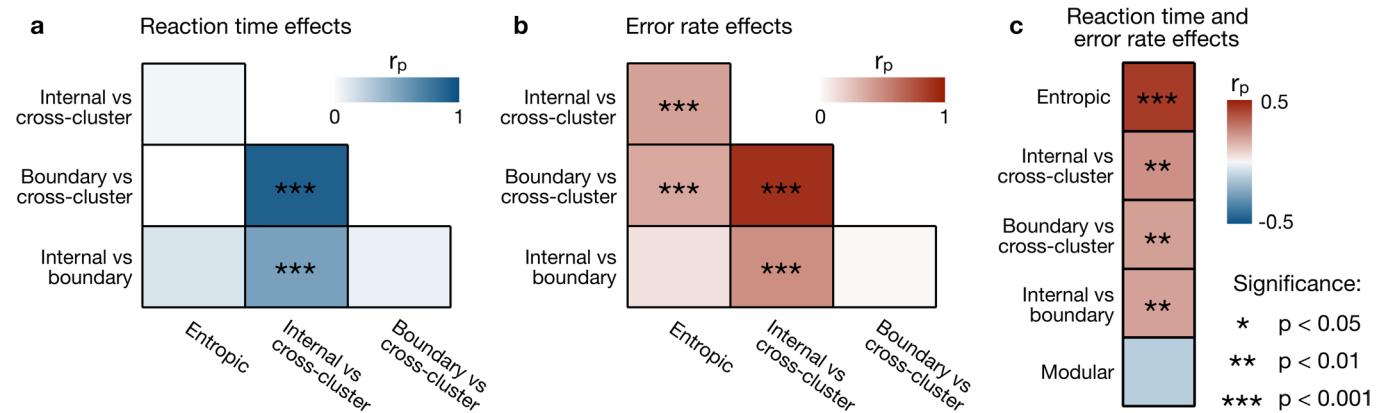
Supplementary information is available for this paper at <https://doi.org/10.1038/s41567-020-0924-7>.

Correspondence and requests for materials should be addressed to D.S.B.

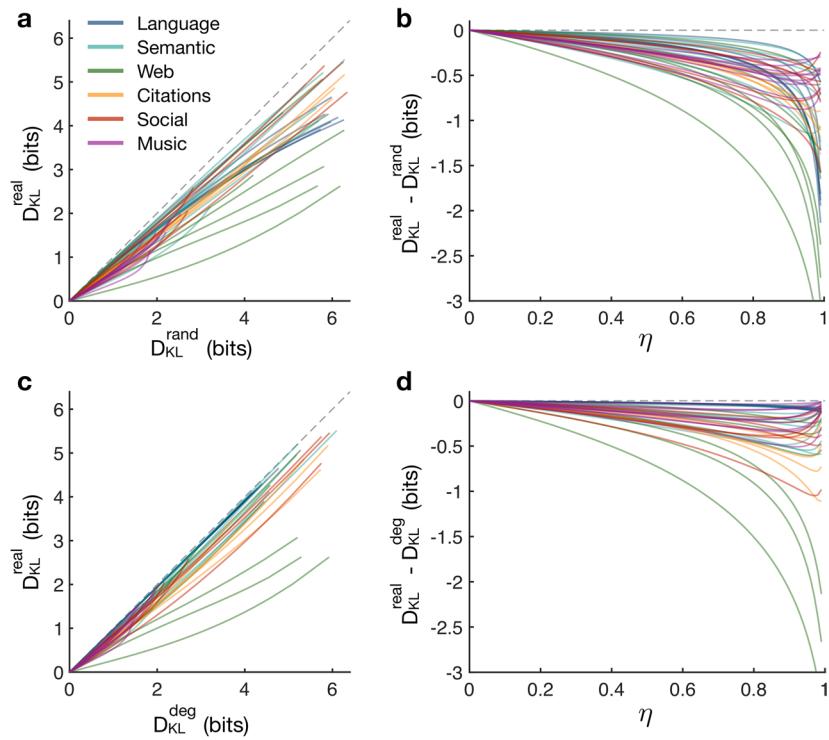
Reprints and permissions information is available at www.nature.com/naturephysics.



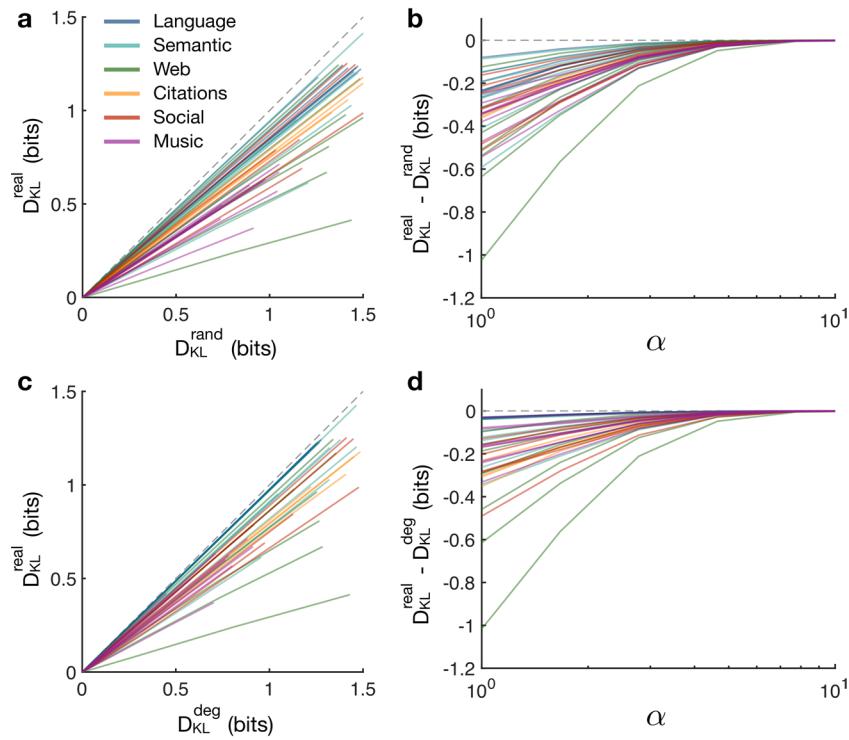
Extended Data Fig. 1 | Distributions of network effects over individual subjects. **a–e**, Distributions over subjects of the different reaction time effects: the entropic effect ($n = 177$), or the increase in reaction times for increasing produced information (**a**); the extended cross-cluster effects ($n = 173$), or the difference in reaction times between internal and cross-cluster transitions (**b**), between boundary and cross-cluster transitions (**c**), and between internal and boundary transitions (**d**) in the modular graph; and the modular effect ($n = 84$), or the difference in reaction times between the modular network and random k -4 networks (**e**). **f–j**, Distributions over subjects of the different effects on error rates: the entropic effect (**f**), the extended cross-cluster effects (**g–i**), and the modular effect (**j**).



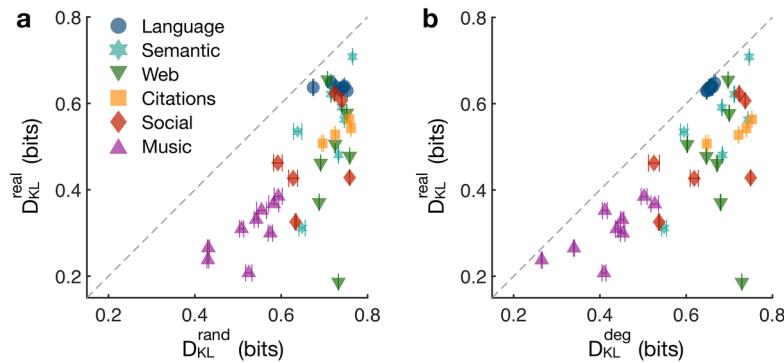
Extended Data Fig. 2 | Correlations between different network effects across subjects. **a**, Pearson correlations between the entropic and extended cross-cluster effects on reaction times ($n = 142$ subjects). **b**, Pearson correlations between the entropic and extended cross-cluster effects on error rates ($n = 142$ subjects). In **a** and **b**, the modular effects on reaction time and error rates are not shown because they were measured in a different population of subjects. **c**, Pearson correlation between the impact on reaction time and error rate for the entropic effect ($n = 177$ subjects), extended cross-cluster effects ($n = 173$ subjects), and the modular effect ($n = 84$ subjects). Statistically significant correlations are indicated by p -values less than 0.001 (**), less than 0.01 (**), and less than 0.05 (*).



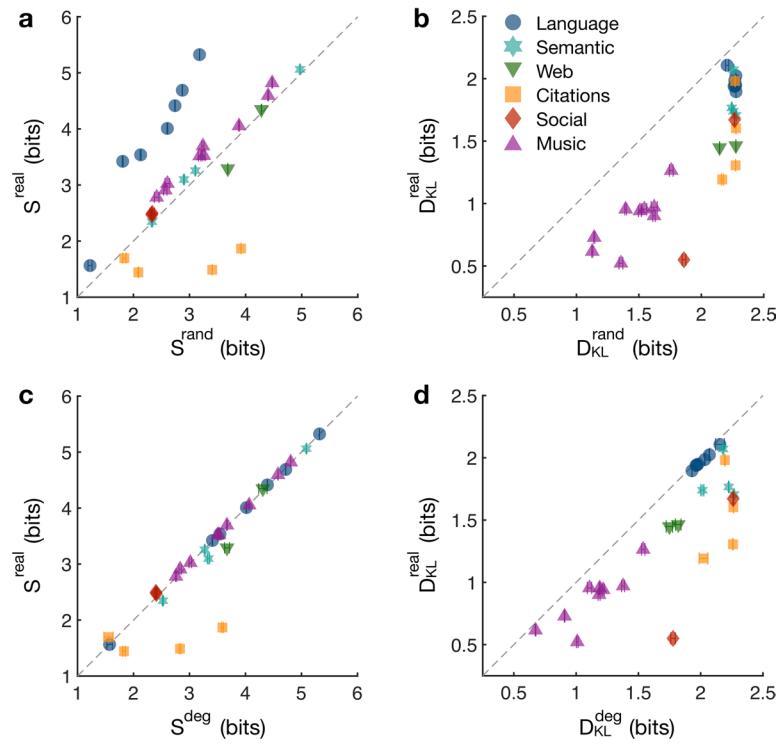
Extended Data Fig. 3 | KL divergence of real networks for different values of η . **a**, KL divergence of fully randomized versions of the real networks listed in Table S12 (D_{KL}^{rand}) compared with the true value (D_{KL}^{real}) as η varies from zero to one. Every real network maintains lower KL divergence than the corresponding randomized network across all values of η . **b**, Difference between the KL divergence of real and fully randomized networks as a function of η . **c**, KL divergence of degree-preserving randomized versions of the real networks (D_{KL}^{deg}) compared with D_{KL}^{real} as η varies from zero to one. The real networks display lower KL divergence than the degree-preserving randomized versions across all values of η . **d**, Difference between the KL divergence of real and degree-preserving randomized networks as a function of η . All networks are undirected, and each line is calculated using one randomization of the corresponding real network.



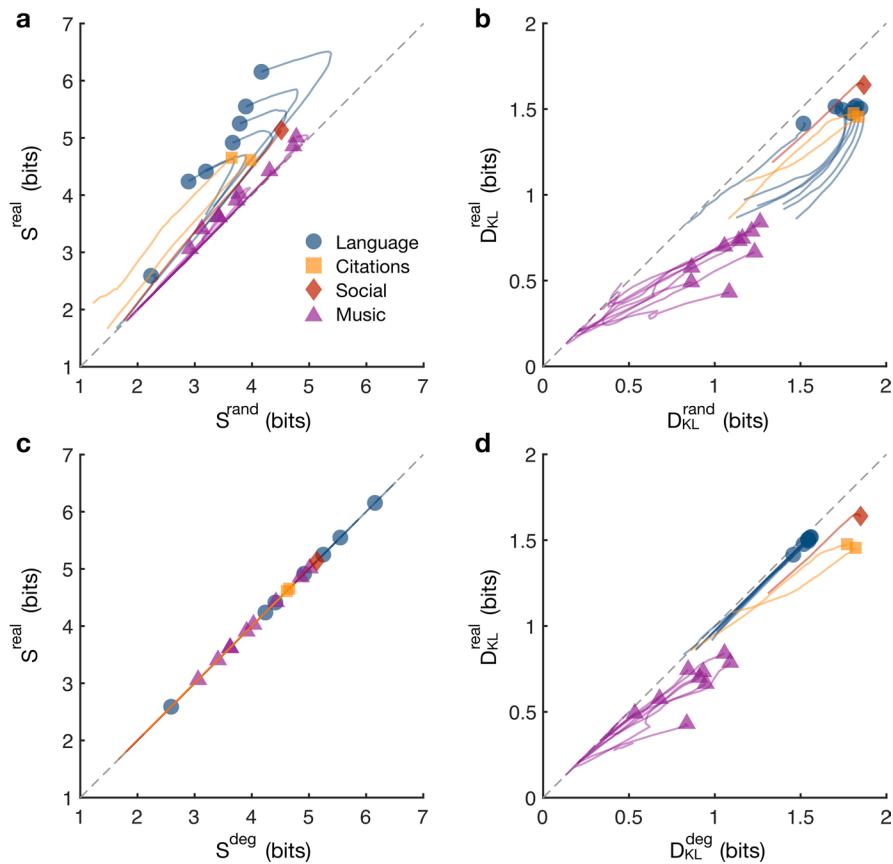
Extended Data Fig. 4 | KL divergence of real networks under the power-law model of human expectations. **a**, KL divergence of fully randomized versions of the real networks listed in Table S12 (D_{KL}^{rand}) compared with the true value (D_{KL}^{real}). Expectations \hat{P} are defined as in Eq. (9) with $f(t) = (t+1)^{-\alpha}$, and we allow α to vary between 1 and 10. The real networks maintain lower KL divergence than the randomized network across all values of α . **b**, Difference between the KL divergence of real and fully randomized networks as a function of α . **c**, KL divergence of degree-preserving randomized versions of the real networks (D_{KL}^{deg}) compared with D_{KL}^{real} as α varies from 1 to 10. The real networks display lower KL divergence than the degree-preserving randomized versions across all values of α . **d**, Difference between the KL divergence of real and degree-preserving randomized networks as a function of α . All networks are undirected, and each line is calculated using one randomization of the corresponding real network.



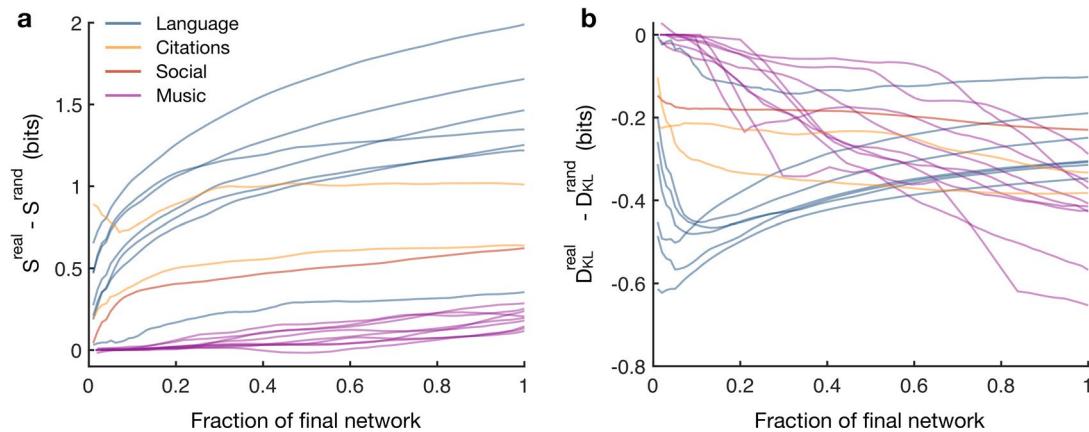
Extended Data Fig. 5 | KL divergence of real networks under the factorial model of human expectations. **a**, KL divergence of fully randomized versions of the real networks listed in Table S12 (D_{KL}^{rand}) compared with the exact value (D_{KL}^{real}). Expectations \hat{P} are defined as in Eq. (9) with $f(t) = 1/t!$. **b**, KL divergence of degree-preserving randomized versions of the real networks (D_{KL}^{deg}) compared with D_{KL}^{real} . In both cases, the real networks maintain lower KL divergence than the randomized versions. Data points and error bars (standard deviations) are estimated from 10 realizations of the randomized networks.



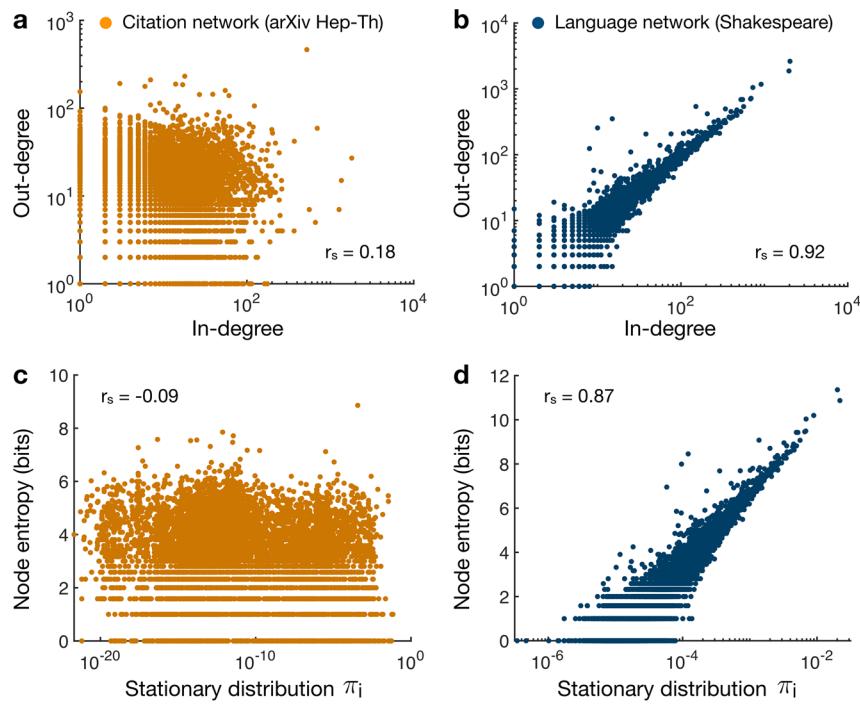
Extended Data Fig. 6 | Entropy and KL divergence of directed versions of real networks. **a**, Entropy of directed versions of the real networks listed in Table S12 (S^{real}) compared with fully randomized versions (S^{rand}). Entropy is calculated directly from Eq. (2) with the stationary distribution π calculated numerically. **b**, KL divergence of directed versions of the real networks ($D_{\text{KL}}^{\text{real}}$) compared with fully randomized versions ($D_{\text{KL}}^{\text{rand}}$). Expectations \hat{P} are defined as in Eq. (10) with η set to the average value 0.80 from our human experiments. **c**, Entropy of randomized versions of directed real networks with in- and out-degrees preserved (S^{deg}) compared with S^{real} . **d**, KL divergence of degree-preserving randomized versions of directed real networks ($D_{\text{KL}}^{\text{deg}}$) compared with $D_{\text{KL}}^{\text{real}}$. Data points and error bars (standard deviations) are estimated from 100 realizations of the randomized networks.



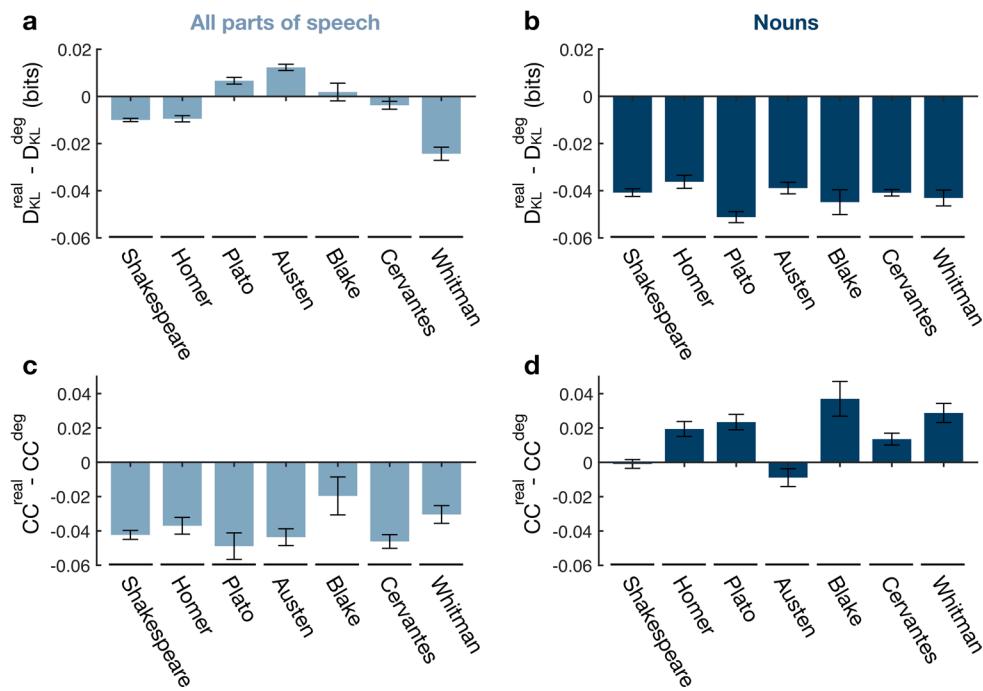
Extended Data Fig. 7 | Entropy and KL divergence of temporally evolving versions of real networks. Entropy of temporally evolving versions of the real networks listed in Table S12 (S^{real}) compared with fully randomized versions (S^{rand}). Each line represents a sequence of growing networks and each symbol represents the final version of the network. **b**, KL divergence of evolving versions of the real networks ($D_{\text{KL}}^{\text{real}}$) compared with fully randomized versions ($D_{\text{KL}}^{\text{rand}}$). Expectations \hat{P} are defined as in Eq. (10) with η set to the average value 0.80 from our human experiments. **c**, Entropy of temporally evolving versions of real networks (S^{real}) compared with degree-preserving randomized versions (S^{deg}). **d**, KL divergence of temporally evolving versions of real networks ($D_{\text{KL}}^{\text{real}}$) compared with degree-preserving randomized versions ($D_{\text{KL}}^{\text{deg}}$). Across all panels, each point along the lines represents an average over five realizations of the randomized networks.



Extended Data Fig. 8 | Evolution of the difference in entropy and KL divergence between real networks and randomized versions. **a**, Difference between the entropy of temporally evolving real networks (S^{real}) and the entropy of fully randomized versions of the same networks (S^{rand}) plotted as a function of the fraction of the final network size. Each line represents a sequence of growing networks that culminates in one of the communication networks studied in the main text. **b**, Difference between the KL divergence of temporally evolving real networks ($D_{\text{KL}}^{\text{real}}$) and that of fully randomized versions ($D_{\text{KL}}^{\text{rand}}$) plotted as a function of the fraction of the final network size. When calculating the KL divergences, the expectations \hat{P} are defined as in Eq. (10) with η set to the average value 0.80 from our human experiments. Across both panels, each point along the lines represents an average over five realizations of the randomized networks.



Extended Data Fig. 9 | Comparison of directed citation and language networks. **a**, Out-degrees $k_i^{\text{out}} = \sum_j G_{ij}$ of nodes in the arXiv Hep-Th citation network compared with the in-degrees $k_i^{\text{in}} = \sum_j G_{ji}$ of the same nodes; we find a weak Spearman's correlation of $r_s = 0.18$. **b**, Out-degrees compared with in-degrees of nodes in the Shakespeare language (noun transition) network; we find a strong correlation $r_s = 0.92$. **c**, Entries in the stationary distribution π_i for different nodes in the citation network compared with the node-level entropy S_i ; we find a weakly negative Spearman's correlation $r_s = -0.09$. **d**, Entries in the stationary distribution compared with node-level entropies in the language network; we find a strong Spearman's correlation $r_s = 0.87$.



Extended Data Fig. 10 | Comparison of all-word transition networks and noun transition networks. **a-b**, Difference between the KL divergence of language (word transition) networks (D_{KL}^{real}) and degree-preserving randomized versions of the same networks (D_{KL}^{deg}). We consider networks of transitions between all words (**a**) and networks of transitions between nouns (**b**). **c-d**, Difference between the average clustering coefficient of language networks (CC^{real}) and degree-preserving randomized versions of the same networks (CC^{deg}), where transitions are considered between all words (**c**) or only nouns (**d**). In all panels, data points and error bars (standard deviations) are estimated from 100 realizations of the randomized networks, and the networks are undirected.

Reporting Summary

Nature Research wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Research policies, see [Authors & Referees](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a Confirmed

- The exact sample size (n) for each experimental group/condition, given as a discrete number and unit of measurement
- A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
- The statistical test(s) used AND whether they are one- or two-sided
Only common tests should be described solely by name; describe more complex techniques in the Methods section.
- A description of all covariates tested
- A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
- A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
- For null hypothesis testing, the test statistic (e.g. F , t , r) with confidence intervals, effect sizes, degrees of freedom and P value noted
Give P values as exact values whenever suitable.
- For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
- For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
- Estimates of effect sizes (e.g. Cohen's d , Pearson's r), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection

Custom code was used to collect data on Amazon's Mechanical Turk. This code has been made publicly available, and can be found at <https://github.com/ariekahn/psiTurk>.

Data analysis

To analyze data, we use built-in functions in MATLAB (R2018a).

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors/reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Research [guidelines for submitting code & software](#) for further information.

Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A list of figures that have associated raw data
- A description of any restrictions on data availability

Source data for Fig. 1, Supplementary Figs. 3-5, and Supplementary Tables 1-11 are provided in Supplementary Data File 1. Source data for Fig. 2 and Supplementary Fig. 1 are provided in Supplementary Data File 2. Source data for the networks in Fig. 3, Table 1, and Supplementary Figs. 6-9 are either publicly available or provided in Supplementary Data File 3 (see Supplementary Table 12 for details).

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

- Life sciences Behavioural & social sciences Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Behavioural & social sciences study design

All studies must disclose on these points even when the disclosure is negative.

Study description	We perform one type of study, wherein we measure participants' reaction times and error rates in serial response experiments.
Research sample	Participants were selected using Amazon's Mechanical Turk, which provides unprecedented access to human behavioral data from participants around the globe. Using this population allowed us to gather large amounts of behavioral data from a multi-national sample of participants. Previously-collected data, also from Amazon's Mechanical Turk, was used in some of our analyses; this data can be found in [Kahn, Ari E., et al., <i>Nat. Hum. Behav.</i> (2018)].
Sampling strategy	Participants signed up for the experiments voluntarily online. We ensured that each participant was unique to preclude the possibility of a participant performing two different experiments. We selected sample sizes based on previously-used protocols described in [Kahn, Ari E., et al., <i>Nat. Hum. Behav.</i> (2018)].
Data collection	The data was collected on Amazon's Mechanical Turk. The data were collected on participants' personal computers with no researchers present.
Timing	The data was collected between 4/11/19 and 4/18/19.
Data exclusions	No subjects were excluded from analyses. We excluded all trials in which subjects responded incorrectly. Finally, we excluded reaction times that were implausible, either three standard deviations from a subject's mean reaction time, below 100 ms, or over 3500 ms.
Non-participation	No participants dropped out or declined participation.
Randomization	Participants were not allocated into experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Antibodies |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Eukaryotic cell lines |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Palaeontology |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Animals and other organisms |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Human research participants |
| <input type="checkbox"/> | <input checked="" type="checkbox"/> Clinical data |

Methods

- | | |
|-------------------------------------|---|
| n/a | Involved in the study |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> ChIP-seq |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> Flow cytometry |
| <input checked="" type="checkbox"/> | <input type="checkbox"/> MRI-based neuroimaging |

Human research participants

Policy information about [studies involving human research participants](#)

Population characteristics	See above.
Recruitment	Participants signed up for the experiments voluntarily on Amazon's Mechanical Turk. Given the simplicity of the experiments, we do not foresee any biases that would reflect the results.
Ethics oversight	All participants provided informed consent in writing and experimental methods were approved by the Institutional Review Board of the University of Pennsylvania.

Note that full information on the approval of the study protocol must also be provided in the manuscript.