PNAS RESEARCH ARTICLE | PSYCHOLOGICAL AND COGNITIVE SCIENCES
NEUROSCIENCE

Medial temporal lobe encodes cognitive maps of real-world social networks

Yi Yang Teoh[a,1], Jae-Young Son[a,1], Alice Xia[a], Apoorva Bhandari[a,2,3], and Oriel FeldmanHall[a,b,2,3]

Affiliations are included on p. 12.

Edited by Uta Frith, University College London, London, United Kingdom; received August 22, 2025; accepted March 6, 2026

Humans routinely solve social problems by navigating densely interconnected networks—gossiping strategically, brokering across cliques, and coordinating with allies. Yet, the neural representations supporting such complex navigation remain undocumented, leaving a crucial gap in our understanding of the mechanisms underlying flexible social cognition. We combine computational modeling with fMRI and a longitudinal measurement of a large, real-world social network (N = 187) to show that the medial temporal lobe encodes an abstract cognitive map of long-range connectivity in the broader network. These neural maps specifically encode the Katz communicability between network members, not merely direct ties, or Euclidean or graph distance. We additionally find that the availability of these maps in the right entorhinal cortex (rEC) and right anterior hippocampus (raHC) supports accurate inferences about information diffusion through the network. Finally, we link neural representation to consequential social change, finding that stronger rEC encoding predicts real-world brokerage that increases community cohesion over time. Together, these findings advance a general mechanism for how the brain supports flexible behavior in complex social environments, closing a major gap between theories of cognitive maps and the demands of social navigation.

cognitive maps | social network | entorhinal cortex | hippocampus

Significance

Navigating everyday social life—from strategic gossiping to building influence—depends on a representation of the hundreds of ties that link people in the community to one another. We uncover the precise format of these neural representations in an ecologically valid context: a large and dynamically changing real-world social network. Entorhinal–hippocampal maps that successfully encode the connectivity between other people in the network (beyond just direct ties) predict better real-world social outcomes, including greater social cohesion within one's community. Results reveal how the neural mechanisms in the medial temporal lobe, known to underpin relational processing in simple grid worlds, are adapted to capture the high-dimensional, non-Euclidean relational complexity of the social world.

To access the rich resources embedded in social networks and flourish (1), people must learn to solve a variety of problems related to navigating complex social environments. From organizing a dinner party where no one feels left out to spreading pertinent gossip about an abusive coworker, our ability to make adaptive choices hinges on whether we can construct a useful mental representation of the social relations within our communities. Different social problems demand different kinds of information—e.g., deep knowledge of social milieu to decide seating arrangements for an intimate dinner, versus inferences about social connections within the broader network that ensure gossip stays contained. Given the sheer number and variety of challenges people face in their social worlds, neural representations of social networks must be structured to enable maximal behavioral flexibility. How exactly does the brain solve this representational problem?

To meaningfully answer this question, we must examine how the brain naturally accomplishes this feat in large and dynamic real-world social networks. Although artificial social networks permit researchers exceptional experimental control during learning (2), it is unknown whether representations of small artificial networks that lack many naturalistic qualities mirror how people spontaneously represent their own social networks in the wild. However, probing naturalistic representations of real-world networks introduces unique challenges: Researchers cannot control how an individual observes their peers' social interactions, and thus have no direct access to what people know about their social networks.

Advancements in computational modeling provide an elegant approach to addressing the challenges of studying natural social networks. Equipped with a minimal set of assumptions about the observations people can realistically make given the constraints of the real world, modeling allows us to synergistically draw on insights from relevant research on cognitive maps in controlled environments (3–8) and apply them to far noisier real-world analogs.

One such insight is the critical role of the hippocampus (HC) and entorhinal cortex (EC) in constructing and maintaining map-like representations of relations between objects, locations, and concepts (9–16). These same regions have also been implicated in the mapping of two-dimensional social spaces of others' power and competence (17, 18), further supporting the idea that the HC and EC play a fundamental role in representing relational information across domains. However, while these prior results suggest that

Author contributions: J.-Y.S., A.X., A.B., and O.F.H. designed the research; J.-Y.S. and A.X. collected the data; Y.Y.T., J.-Y.S., A.X., and A.B. developed the analyses; Y.Y.T. and J.-Y.S. analyzed the data; A.B. and O.F.H. acquired the funding; A.B. and O.F.H. provided supervision; and Y.Y.T., J.-Y.S., A.B., and O.F.H. wrote the paper.

The authors declare no competing interest.

This article is a PNAS Direct Submission.

[1]Y.Y.T. and J.Y.S. contributed equally to this work.

[2]A.B. and O.F.H. contributed equally to this work.

[3]To whom correspondence may be addressed. Email: apoorva_bhandari@brown.edu or oriel.feldmanhall@brown.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas.2523345123/-/DCSupplemental.

Published April 14, 2026.

similar neural computations in the HC and EC may support representations of our own large and complex social networks, no prior work has empirically demonstrated the encoding of real-world, non-Euclidean social spaces in these regions, or characterized their content and format.

What relational information might these regions encode about social networks? One intuitive possibility is that people try to memorize a complete and veridical map of all the relationships they encounter in their social network. However, when we consider the size of real-world social networks, the immense cognitive demands of constructing and deploying such representations would render them highly inefficient (19, 20). Not only does brute-force memorization of all observed relations present a formidable cognitive challenge (21, 22), but information about others' social relations is also not always apparent or directly observable (23). Instead, recent work finds that people construct "fuzzy" mental representations of others' social relations that trade off specific information about individual ties between people in favor of coarsely clustering them based on shared connections (21, 22, 24–26). These abstract representations emerge from a process of "chaining" knowledge about pairwise social relations to uncover the deeper structural connection between people based on their multistep relations (21, 26). For example, if we know that Jane and Chris are friends, and that Chris and Zach are friends, then we might infer that Jane and Zach are also likely to be friends by abstracting over the known pairwise friendships and identifying the indirect connection between them (27, 28). By leveraging this simple mechanism of multi-step abstraction to cluster structurally connected members of the network together, people can adeptly navigate a wide range of social situations—from generating inferences about unobserved but likely relationships within a network (21, 22), to tracking how gossip might spread across dense ties within a community (25).

Given these considerations, it is unlikely that people construct a complete and veridical map of all social relationships they encounter. We instead hypothesize that people's neural representations of their social networks likely reflect abstractions over whatever relations they have observed. Such abstraction would be consistent with research in the nonsocial domain in which the anterior HC and the EC appear to encode abstract representations of task environments, with the HC maintaining representations of multistep connections between states in reinforcement learning tasks (6, 7), and the EC abstracting over these representations to encode more latent structural properties of the state-spaces (3, 5, 8, 29). However, whether the HC and/or EC encodes abstract representations of relational ties in the real world to support adaptive social navigation, remains wholly unknown.

It is also unclear what specific format these representations should take. Competing representational formats prioritizing different theoretical considerations have been proposed. On one hand, because each person maintains multiple relationships with many others, we might expect that an abstract representation of a social network tracks the absolute multistep connectivity of each person to everyone else, independent of their other connections within the network. Such a representational format—captured by Katz communicability (Katz) (30–33)—is particularly well suited to the task of anticipating the natural dynamics of information flowing through the network (25). Gossip, for example, is often simultaneously transmitted across many social ties at once (34–37), and thus demands a representation that accommodates variations in the overall connectedness of each person. On the other hand, representations of relationships in the network can also be structured in such a way that they quantify each person's relative connectivity to all others in the network (21, 22). This representational format—known as the Successor Representation (SR) (38)—tracks the relative likelihood of people sharing multistep connections with others in the network, which resembles the representations often used for navigating other types of environments (6, 20, 38, 39).

That distinct representational formats prioritize the statistics of different modes of network dynamics (33) begs the question of how people's representations of their social network reflect the demands of navigating said network. For example, in spatial environments, the SR reflects the statistics of the position occupied by an agent traversing sequentially through different locations, one path at a time (8). Unlike Katz, which reflects the statistics of simultaneous diffusion across the network common in social transmission, sequential traversals reflected in the SR forcibly normalize the strength of concurrent connections in the environment. Using the SR would limit tracking of social transmission to cases where it unfolds sequentially—e.g., a note being passed from one person to another over time. This also means that the SR discounts the possibility that one can maintain just as strong of a relationship with five friends as they can if they only had one friend. Relatively little work has been done to verify these distinctive assumptions of the SR within the context of a social network (21, 22, c.f. 25). Thus, although the SR's assumptions do not seem to align with the properties of social networks and phenomena, whether people's representations reflect these assumptions remains less clear. Should we find evidence of a SR-like representation of real-world social networks, it could reflect a domain-general constraint on how highly complex relational information is encoded.

To close these fundamental gaps in our understanding of whether social networks are naturally—and spontaneously—represented in the HC and EC in order to scaffold social navigation, we leveraged the unique context of first-year undergraduates leaving home and moving to college. This allowed us to examine the sociocognitive maps people construct when embedded in a newly emerging and rapidly evolving social network. We longitudinally tracked emerging friendships when students first arrived on campus (N = 196), to the stabilization of the network by the end of the academic year. We also queried each person's representation of their network halfway through the first semester, once students established some initial friendships. We scanned a subset of these participants (N = 43) in order to interrogate whether learning about other people's social connections in the wild (no learning occurred in the laboratory) spontaneously elicits representations of the real-world social network in the HC and EC in the absence of an explicit task. By developing and comparing a suite of computational models, we characterize the precise format in which the brain represents social relations in this naturalistic, ecologically valid, and personally relevant social network. Finally, we test whether these neurocognitive maps predict the ability to solve various social navigation problems, from asking whether these maps can track how information flows across the entire network to probing whether they enable the brokerage of friendships to foster greater social cohesion within people's immediate social communities.

Results

People Build Imprecise Cognitive Maps of Their Real-World Social Network. We recruited 196 first-year undergraduates from three dormitories at Brown University and asked each participant to identify their friends among the other 195 participants multiple times throughout the year (Fig. 1*A*; see *Materials and Methods* and *SI Appendix*, Note S1). Drawing on techniques from social network science (40), we then combined these responses across participants

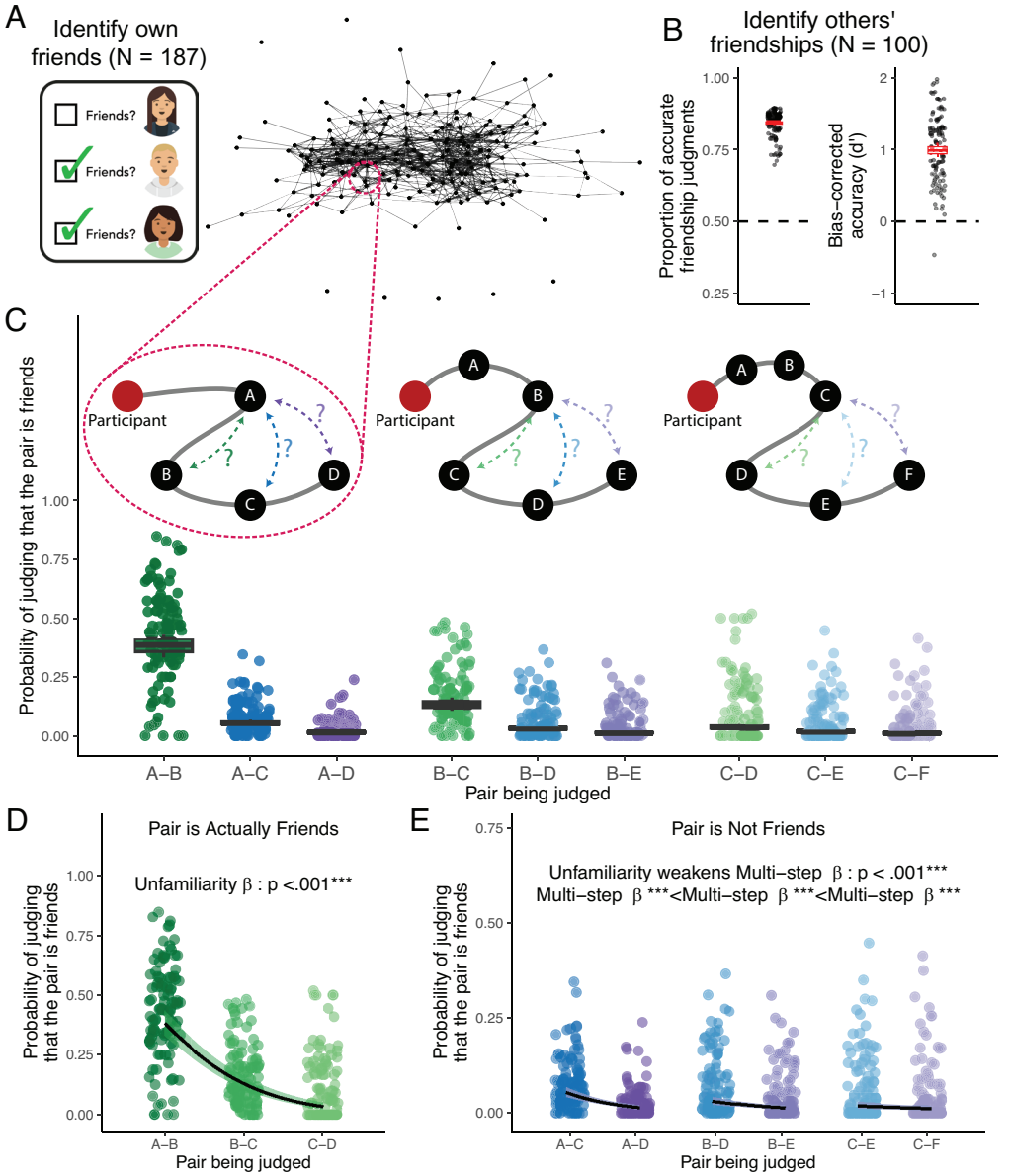

Fig. 1. Friendship knowledge. (*A*) The friendship network constructed from mutually identified friendships. In the graph (*Right*), each node represents a network member who reported their own friendships in the Friendship Identification Task (*Left*). Each edge between two nodes indicates a mutual friendship. Icons: getavataaars.com. (*B*) Accuracy of friendship judgments. The two panels indicate (*Left*) the proportion of accurate choices people made and (*Right*) the bias-corrected accuracy, d-prime, which accounts for the relatively low base-rate of friendships within the network out of all possible dyadic relations that could exist. Each point represents a single participant's data. The central line, upper and lower bounds, and whiskers of the boxplot indicate the estimated mean, std error, and 95% CI from the general linear model. Dashed lines correspond to chance for each metric. (*C*) Friendship judgments reflect both direct and inferred knowledge. Each dot represents the proportion of trials on which a participant indicated that a pair of network members were friends in the Network Knowledge Task, conditional on whether the pair was actually friends (green), separated by a mutual friend (blue: friends-of-friends), or separated by three steps (purple: friends-of-friends-of-friends). These choices were further faceted out by participants' familiarity with the pair as indicated by the visual schematics above the data plots: whether the participant was (*Left*) personally friends with one of the members of the pair, (*Middle*) separated by a mutual friend, or (*Right*) separated by three steps from the pair (i.e., one of the pair's members was a friend-of-a-friend-of-a-friend for the participant). Not all data are plotted here as there were very few trials in which the degree of separation—either within the pair or between the participant and the pair—exceeded three (6.45%). The central, line, upper and lower bounds, and whiskers of the boxplot indicate the estimated mean, within-participant std error, and 95% CI of responses from a logistic mixed-effects regression fitted to all data respectively. To provide a visual guide for the x-axis, which combines both the dyad's distance from each other and their distance from the participant, simplified network schematics are illustrated above the plots. Consider an example not explicitly depicted where the participant is also friends with another network member, X, who is not friends with A-F. If participants are asked to judge if A & X are friends, this would be treated as equivalent to judging if A & C are friends, since A is a personal friend of the participant, and A & X are separated by a mutual friend (i.e., the participant). (*D*) Friendship knowledge depends on familiarity with the pair. This panel replots participants' choices in the Network Knowledge Task in (C), but only for pairs of targets and probes that were actually friends (i.e., hits), highlighting that people are less able to identify friends with whom they are less familiar. Each point represents the proportion of trials in which a participant indicated that the pair was friends. The black line and shaded region indicate the estimated mean and within-participant 95% CI of responses from the fitted logistic mixed-effects regression. (*E*) Erroneous multistep inferences. This panel reproduces participants' choices in the Network Knowledge Task in (C), but only for pairs of targets and probes that were not actually friends (i.e., false alarms), highlighting that people are more likely to (incorrectly) infer friendship between dyads who are separated by two (blue) vs. three steps (purple). However, these multistep inferences are modulated by the participant's familiarity with the pair, such that greater unfamiliarity decreases the likelihood of making incorrect multistep inferences (unfamiliarity x multistep interaction $P < 0.001$). Each point represents the proportion of trials in which a participant indicated that the pair was friends. The black line and shaded region indicate the estimated mean and within-participant 95% CI of responses from the fitted logistic mixed-effects regression.***$P < 0.001$

to derive the "ground-truth" structure of the friendship network based on pairs of network members that mutually identified each other as friends. A friendship is recorded if both A and B mutually report being friends. After the second measurement of the network, approximately one month into the academic year, when students began to establish some initial friendships (N = 187), we recruited a subset of these network members (N = 100) to complete a separate laboratory task probing participants' knowledge of others' relationships in their social network. In this Network Knowledge Task, participants were iteratively presented with the name and photograph of one network member (target) alongside a list of names and photographs of other network members (probes) and asked to judge whether each pair was friends.

Importantly, we note that participants were never explicitly asked to infer friendships based on the structure of the broader network, meaning any multistep inferences were spontaneously generated. The list of targets and probes included a subset of members varying in distance from the participant (e.g., friends, friends of friends, and beyond; *SI Appendix*, Note S1). To then assess the accuracy of participants' network knowledge, we compare these subjective judgments to the ground-truth network based on reciprocal friendships.

We find that participants were able to identify whether other network members were friends with each other at above-chance accuracy [Chance = 0.50; Proportion correct responses: M = 0.843, 95% CI = (0.834, 0.850), z = 53.719, $P < 0.001$, Fig. 1*B*]. This was true even when using indices of accuracy that account for the low-base rate of friendships (i.e., only 14.5% of the dyads presented to participants were actually friends on average): d-prime confirms that people possess actual knowledge of friendships and were not simply responding randomly according to base rates [d-prime: M = 0.991, 95% CI = (0.894, 1.089), t(99) = 20.203, $P < 0.001$, compared to chance-level responding of zero; Fig. 1*B*].

However, as expected, participants also made systematic errors that demonstrate 1) limitations in their knowledge about the network, and 2) the use of multistep abstraction to make inferences about the existence of others' friendships (*SI Appendix*, Table S1). To illustrate these patterns of errors, we separately examined when participants missed actual friendships in the network (hits) and when they mistakenly identified a friendship where none existed (false alarms). We also examined how participants' familiarity with the pair of network members influenced the hit and false alarm rate. In this analysis, we characterized each dyad that the participants judged based on the minimum graph distance between the pair of network members and the participant's minimum graph distance to the dyad as their familiarity with the pair. For instance, if the dyad being considered, Jane & Zach, are friends-of-friends-of-friends, and the participant, Alex, is Jane's friend but not Zach's friend, the minimum graph distance between Jane & Zach is three, and Alex's minimum distance to the dyad is one (Fig. 1*C*: pair A–D, where Alex is the participant, Jane is represented by A, and Zach is represented by D).

First, even in the best-case scenario where a participant is friends with one of the people in the dyad (Fig. 1*C*, pair A–B in dark green), the participant only accurately identifies ~40% of true friendships on average [hit rate: M = 0.382, 95% CI = (0.330, 0.433)] —a far cry from having complete knowledge of the network's structure. This accuracy further degrades as a function of the participant's familiarity with the pair being evaluated [unfamiliarity $\beta = -1.398$, 95% CI = (−1.588, −1.207), z = −14.383, $P < 0.001$; pair A–B vs pair B–C vs pair C–D, Fig. 1*D*]. If the participant is not directly friends with anyone in the dyad, but is instead separated by a mutual friend, the participant becomes less accurate [i.e., B–C: M = 0.132, 95% CI = (0.104, 0.160)], which decreases even further when separated by three-degrees [C–D: M = 0.036, 95% CI = (0.023, 0.050)].

Second, participants also erroneously endorse nonexistent friendships (false alarms) in systematic ways that are consistent with a mental representation of the network that goes beyond individual friendships to cluster network members based on their multistep connectivity. When a participant evaluates whether their friend is friends with someone else—but the two are not in fact friends, the participant is more likely to incorrectly infer a friendship if the pair share mutual friends [triadic closure (28): A–C: M = 0.054, 95% CI = (0.042, 0.066)] compared to if the pair are separated by three degrees [pair A–D: M = 0.026, 95% CI = (0.019, 0.033); A–C vs A-D: Multistep $\beta = -1.354$, 95% CI = (−1.618, −1.091), z = −10.067, $P < 0.001$, Fig. 1*E*]. These erroneous friendship endorsements reveal that participants spontaneously rely on multistep connectivity within the network to infer who is friends with whom, despite its detrimental effects for accurately reporting friendships.

This multistep abstraction persists even if the participant is herself not directly friends with one person in the pair but is separated by mutual friends [B-D vs B-E: multistep $\beta = -0.920$, 95% CI = (−1.094, −0.746), z = −10.341, $P < 0.001$] or three-degrees of friendship [C–E vs C–F: Multistep $\beta = -0.486$, 95% CI = (−0.717, −0.255), z = −4.124, $P < 0.001$]—despite the overall frequency of these inferences decreasing as a function of being unfamiliar with the pair [difference in Multistep β with increasing unfamiliarity = 0.434, 95% CI = (0.258, 0.610), z = 4.834, $P < 0.001$]. Therefore, people's explicit judgments about the structure of the network are behaviorally consistent with the notion that 1) their ability to acquire information about all the friendships in the network is limited, and 2) they use multistep abstraction to make inferences about potential relationships never observed.

Abstract Cognitive Maps of Network Ties Independently Encode Concurrent Relations. To compare different formats of network representation that could potentially support these behavioral patterns, we turn to computational modeling, which allows us to directly characterize the underlying cognitive representations. We consider four possible representational models of the social network. First, as a baseline for comparison, we consider an unlikely (null) model where people's representations of the network comprise perfect knowledge about all pairwise mutual friendships, such that errors in judgments simply result from noisy application of this knowledge. A second model tests our first main hypothesis that people's ability to observe friendships is limited by where they sit in the network (i.e., the limited observation model). This model produces idiosyncratic representations of the social network based on each person's specific position within it, assuming that people are most likely to observe social interactions occurring within their immediate "social circle" and are less likely to observe more distal relations (41). We estimate a parameter $\omega \in [0,1]$ to quantify how strongly observations are affected by this distance: When $\omega \to 0$, an individual is estimated to possess only direct information about their own friendships, and as $\omega \to 1$, individuals are estimated to have access to information about all friendships in the network. To test our second main hypothesis that people abstract over these limited observations, we consider two additional candidate models of multistep abstraction based on prior work: Katz communicability and the Successor Representation (SR). Both models similarly generate abstract representations of associations between every pair of network members as a weighted sum of all the multistep relations between them, weighing each multistep relation proportionally based on the number of steps it comprises

by a factor of $\gamma \in [0,1)$ (8, 20–22, 25, 31–33, 39). A direct relation is weighted by γ, while a two-step relation is weighted by γ^2, etc. When $\gamma \to 0$, the model reflects direct observations of pairwise social relations. As $\gamma \to 1$, the representation becomes increasingly abstract, reflecting greater integration of the longer-range connections between a pair of individuals. Despite these algorithmic similarities, these models make distinctive assumptions about the precise representational format of relational information that in turn generate systematically divergent predictions about the inferred connectivity between individuals (*SI Appendix*, Note S2). In prioritizing the encoding of relative over absolute connectivity in the network, the SR, unlike Katz, includes a normalization step that results in unique penalties for individuals with more friends: Popular individuals are assumed to have weaker individual connections to others (Fig. 2*A*).

At first blush, computational modeling results suggest that, among the candidate models, the Katz model best captures participants' behavior [average Pearson correlation between model-predicted and observed judgments: mean r = 0.857, 95% CI = (0.815, 0.898), t(99) = 40.857, $P < 0.001$; Fig. 2*B*]. Its performance is closely followed by the limited observation model [mean r = 0.841, 95% CI = (0.795, 0.886), t(99) = 36.875, $P < 0.001$] and then the SR model [mean r = 0.834, 95% CI = (0.790, 0.879), t(99) = 37.218, $P < 0.001$]. All three models handily outperform the perfect observation (null) model [mean r = 0.627, 95% CI = (0.590, 0.664), t(99) = 33.362, $P < 0.001$]. This is confirmed by formal model comparisons that consistently identify Katz as the best model of participants' behavior: The Katz model is not only the best model of participants' data in aggregate but also the best-fitting model for the majority of participants (average participant-level model weight (42) based on WAIC (43–45): w(WAIC) = 0.605; protected exceedance probability (46): pxp > 0.999; overall sample w(WAIC) > 0.999; Fig. 2*C*). In other words, our results strongly suggest that people's representations of their social network result from multistep abstraction over limited observations. Moreover, the fact that Katz also outperforms the SR in explaining participants' inferences about relationships (see *SI Appendix*, Note S3 for further evidence from out-of-sample predictions) confirms our a priori predictions that people's representations of their social network are structured by default to encode each dyadic relation independently, reflecting the demands of flexible social navigation which includes the ability to track simultaneous social transmission.

To further demonstrate the adaptive utility of abstraction in generating accurate inferences when direct observation is limited, even if it occasionally leads to errors, we examine whether people who engage in more abstraction exhibit more accurate knowledge of friendships in their network. In the best-fitting Katz model, there are two parameters that dictate the accuracy of friendship judgments: being able to directly observe far-away social relationships (ω), and the amount of abstraction over these direct observations (γ). Unsurprisingly, participants estimated to have higher ω were more accurate in their friendship judgments [$\beta = 0.549$, 95% CI = (0.072, 1.025), t(95) = 2.286, $P = 0.024$]. Critically, over and above the influence of direct observation, accuracy in friendship judgments was also strongly associated with how much a participant abstracts over those observations (γ). This relationship exhibits an inverse u-shape pattern [quadratic term: $\beta = -8.723$, 95% CI = (−14.957, −2.489), t(95) = −2.778, $P = 0.007$; Fig. 2*D* and *SI Appendix*, Table S2], such that a moderate level of abstraction yields the greatest benefits for accurately identifying friendships (vertex of the parabola: $\gamma = 0.682$), and engaging in any more or less abstraction than this optimum yields lower accuracy. These modeling results support our key hypothesis that people are limited in their ability to construct a high-fidelity representation of their social network, but they can "fill in the gaps" using an appropriate multistep abstraction mechanism to infer the existence of unobserved social ties.

Entorhinal Cortex Encodes an Abstract Cognitive Map of One's Social Network. A fundamental benefit of constructing abstract cognitive maps is that they can be encoded in memory during learning and efficiently retrieved when needed for subsequent choices (19, 20). However, it remains an open question where in the brain these cognitive maps of social networks are stably encoded. We recruited a subset of participants who completed the Network Knowledge Task to return for a separate functional MRI session (fMRI; N = 43). Participants were presented with photographs of network members, one at a time, and asked to press a button whenever an upside-down image (i.e., a stock photo of a stranger who was not a member of the network) was presented, allowing us to examine whether the spontaneous neural activation of one's peers naturally comprises information about their social relations in the network (2, 47–49). Using Representational Similarity Analysis (RSA), we estimated how similarly (or dissimilarly) the brain encodes each network member based on the cross-validated Mahalanobis distance between patterns of neural activity spontaneously evoked by the member's photograph (50, 51). We computed matrices of pairwise neural similarity in six regions of interest (ROIs) across the medial temporal lobe based on their known role in encoding cognitive maps: left and right anterior hippocampus (HC), left and right posterior hippocampus, and left and right entorhinal cortex (EC; see *Materials and Methods*). We then examined whether (and how strongly) these regions encode participants' model-estimated (Katz) representation of the network. Each participant's Katz representation comprises a unique matrix of communicability values reflecting their perception of the overall connectivity between each network member (row) and all other network members (column). The value in each cell of this matrix reflects the integrated sum of the participant's direct knowledge of a dyad's pairwise friendship and their inferences about the dyad's multistep connections, controlled by the participant-specific estimates of ω and γ respectively. We first tested whether greater connectivity between two network members—expressed by higher Katz communicability values—predicts more similar patterns of activity in the HC and/or EC (Fig. 2*E* and *SI Appendix*, Table S3).

We find that the right EC (rEC) encodes network connectivity. The more connected a pair of network members are perceived to be across multistep relations—i.e., have greater Katz communicability between them—the more they evoked dissimilar patterns of neural activity in rEC [rEC ~ Katz $\beta = -0.939$, 95% CI = (−1.599, −0.278), t(30.407) = −2.901, $P = 0.007$ uncorrected, Bonferroni-corrected at $\alpha = 0.05$ for 6 ROIs, $P < 0.008$; Fig. 2*F*]. Although this dissimilarity code may seem counterintuitive at first glance, the fact that rEC encodes relevant information about multistep connectivity illustrates its utility as a functional map of the social network's structure, which we verify further in subsequent analyses. In the right anterior HC, we only observed marginal effects of connectivity between network members on neural pattern similarity [raHC ~ Katz $\beta = 0.804$, 95% CI = (−0.106, 1.714), t(29.174) = 1.806, $P = 0.081$ uncorrected; Fig. 2*G*]. There was no evidence that relational information is encoded in the right posterior HC or any of the left-lateralized ROIs (Ps ≥ 0.351), or the control visual regions (left and right V1: Ps ≥ 0.333). Supplemental analyses further controlling for visual similarities in the images (*SI Appendix*, Table S4) and other network properties (*SI Appendix*, Note S4) confirm the robustness of our effects.

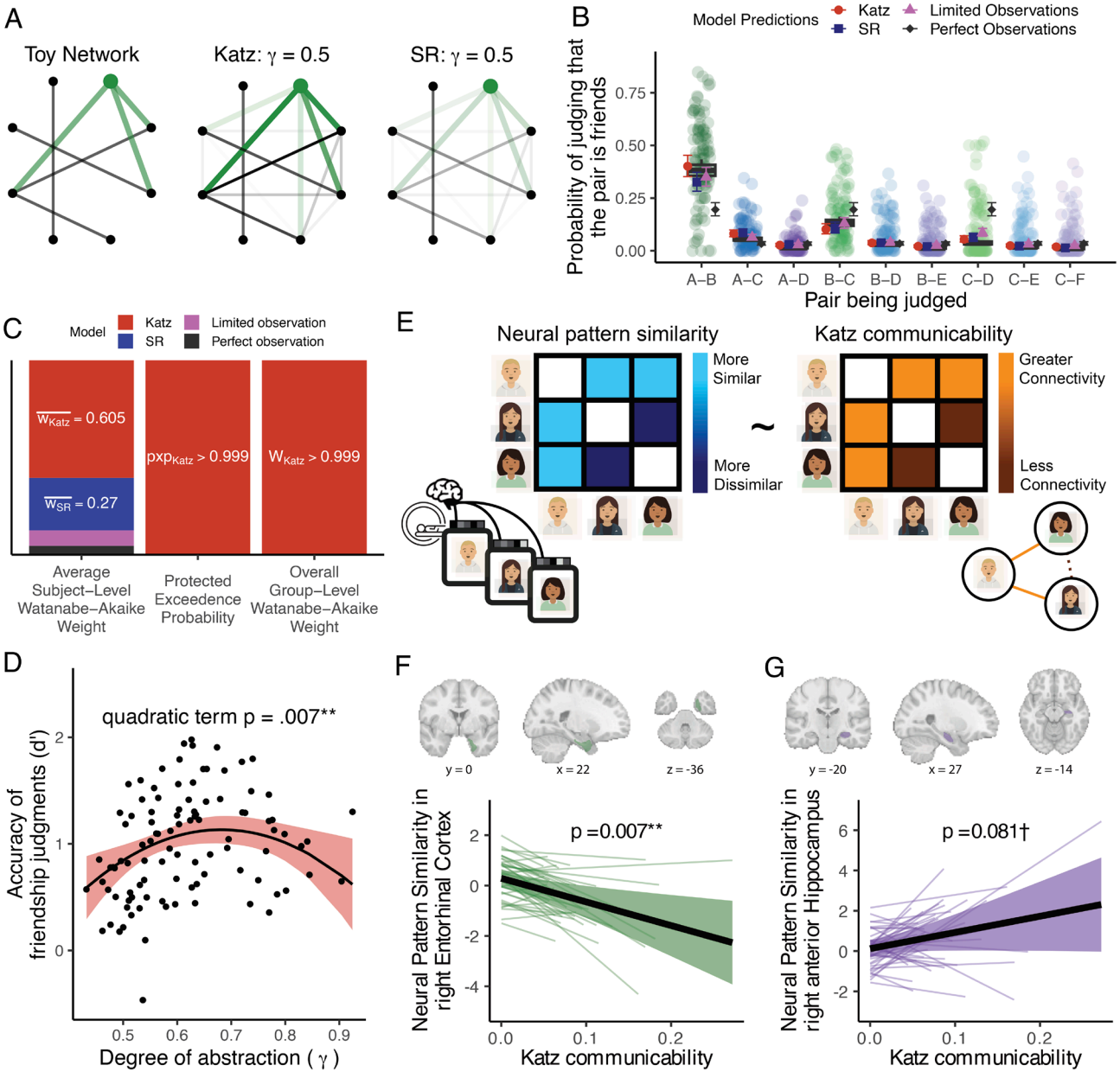

Fig. 2. Computational models of network representation. (*A*) Divergent representations of a toy network from Katz and SR. Nodes represent each of eight members in a toy network. The green node and edges identify the most influential member of the toy network (by eigenvector centrality) and their connections. The *Left* panel illustrates the true toy network composed of edges between nodes signifying a mutual friendship. The *Middle* panel and *Right* panel illustrate representations of the network generated from a Katz and SR model respectively with unlimited observations of friendships at a moderate level of abstraction (γ = 0.5). The color of the edges in these panels indicates the strength of the multistep connection between nodes, with more saturated colors indicating a stronger multistep connection. To facilitate comparison across the different scales of edge weights, we normalize values with reference to the single tie between the two nodes (positioned north–north–west and south–south–west) that are unconnected to the rest of the network. (*B*) Model fits. Points and error bars projected over scatterplots and boxplots of participant data (also in Fig. 1) indicate the estimated mean and 95% CI based on the SE of model predictions averaged across participants. Shape and color of the point and error bars indicate whether the predictions come from the Katz, SR, or models that veridically represent Limited or Perfect Observations. (*C*) Model comparison. Height of stacked bars indicates the relative evidence for each of the four candidate models based on the average participant-level model weight based on the Watanabe–Akaike information criterion, protected exceedance probability (pxp) across participants, and the overall group-level Watanabe–Akaike weight. Labels are omitted when model indices < 0.1 [mean w(limited) = 0.080, mean w(perfect) = 0.046]. (*D*) The benefits of abstraction. The curvilinear relationship between the estimated degree of abstraction (γ) from the best-fitting Katz model and accuracy in the friendship judgment task (d-prime) suggests an optimal degree of abstraction for correctly identifying friendships in the network. Each point indicates a participant. The line and shaded region indicate the estimated mean and 95% CI of predictions from a linear model permitting a quadratic relationship between abstraction and accuracy. (*E*) Neural representational similarity analysis. To test whether the brain encodes an abstract representation of social network structure, we use Katz communicability between network members to predict neural patterns computed from activity evoked during passive viewing of network members' photographs during fMRI. Icons: Flaticon.com & getavataaars.com (*F*) Right entorhinal cortex (rEC) pattern similarity. Neural pattern dissimilarity in rEC is significantly predicted by Katz communicability. (*G*) Right anterior hippocampus (raHC) pattern similarity. Neural pattern similarity in raHC is marginally predicted by Katz. Each thin colored line in B-C indicates the best-fit line for each individual participant from a mixed-effects linear model fitted hierarchically to the data. Solid black lines and colored shaded regions indicate the mean and 95% CI of model-estimated predictions for the group, respectively. †$P < 0.1$, **$P < 0.01$ (uncorrected).

To unpack these neural results, we conducted two additional follow-up analyses to elaborate on the specific nature and content of the observed representations. Our results suggest that the rEC encodes a cognitive map of multistep social connectivity using a dissimilarity code. However, because familiarity with a network member likely reflects their relevance for participants' social lives

and facilitates participants' ability to acquire more information about them, a potential confound is that rEC pattern dissimilarity simply reflects participants' familiarity with network members: Participants might represent network members closest to themselves more distinctly from each other and from other more distant network members (52–55). To rule this out, we examined whether rEC pattern similarity between network members changes as a function of participants' familiarity with the most proximal member of the dyad. We find no evidence that rEC patterns exhibit more differentiated representations of participants' friends compared to more distant network members [unfamiliarity β = 0.172, 95% CI = (−0.039, 0.383), t(41.814) = 1.642, P = 0.108, *SI Appendix*, Table S5]. Furthermore, simultaneous inclusion of unfamiliarity *and* Katz communicability in a regression model instead confirms that multistep connectivity robustly predicts representational dissimilarity in rEC [Katz β = −0.778, 95% CI = (−1.401, −0.155), t(32.203) = −2.542, P = 0.016, *SI Appendix*, Table S5]. Thus, these analyses confirm that the dissimilarity code in rEC cannot be attributed to the pattern separation of familiar network members but rather contains specific pairwise relational information consistent with the encoding of a social cognitive map.

Second, because model-estimated Katz communicability values reflect the integration of both inferred multistep relations and direct observations of pairwise ties, it is unclear whether neural encoding of a Katz representation merely reflects the memorization of observed ties, or if these neural processes integrate and encode inferences of unobserved relations through multistep abstraction. To probe whether the abstracted inferences are indeed encoded in the brain, we conducted follow-up analyses that control for the possible observations the participant could make within our computational model (*SI Appendix*, Table S6). We find that Katz-encoding in the rEC remains significant [Katz β = −1.194, 95% CI = (−2.152, −0.237), t(34.278) = −2.534, P = 0.016], while the marginal association between Katz and neural pattern similarity in raHC is abolished [Katz β = 0.950, 95% CI = (−0.895, 2.796), t(15.815) = 1.092, P = 0.291]. In combination with additional supplementary analyses (*SI Appendix*, Table S7-S8), these findings establish the key role of rEC in robustly encoding an abstracted map of multistep connections within one's social network, while also suggesting that the raHC might be encoding a more veridical representation of observed social ties.

Neural Encoding of Relational Maps for Tracking Information Flow. Although the rEC appears to encode relational maps of one's social network, what is the functional utility of these maps for solving navigation problems in the network? One formidable social navigation problem is understanding how information flows across chains of social ties. If an individual understands how information might traverse across the network, it enables several adaptive behaviors, from strategically spreading gossip that can sabotage relationships without getting caught, to enhancing one's reputation by acting generously toward well-connected members. Successfully deploying these strategic behaviors requires an intricate understanding of how the litany of social ties enables information to spread. To test whether sociocognitive maps in rEC and raHC facilitate predictions about information flow, we probed participants' inferences about information flow in a separate task conducted independently from the fMRI session. In this Information Flow Task, participants were instructed that a network member (the source) has hypothetically disclosed some news about themselves and then asked to judge how likely it would be for another network member (the target) to hear this news. Accurate inference in this task requires knowledge of whether, and how, two individuals are connected, and ideally, knowledge of every possible connection between them—as news can travel on any number of paths. Given this, we test the hypothesis that neural maps tracking multistep connectivity between network members would be valuable for tracking information flow. To do this, we investigate whether dissimilarity in participants' spontaneous neural representations of the source and target in rEC and their similarity in raHC predict their judgments of whether information would flow between them (Fig. 3*A*).

Consistent with this prediction (and our finding that Katz communicability is represented using a *dissimilarity* code in rEC), greater dissimilarity between rEC activity evoked by the source and target is associated with inferences that information is more likely to flow between them [β = −0.225, 95% CI = (−0.381, −0.070), z = −2.843, P = 0.004; Fig. 3*B* and *SI Appendix*, Table S9]. Additionally, consistent with our results suggesting that Katz communicability may be represented using a *similarity* code in raHC, greater similarity between neural patterns evoked by the source and target in the raHC is also associated with inferences that information flows between them [β = 0.178, 95% CI = (0.038, 0.318), z = 2.499, P = 0.012; Fig. 3*C*]. Furthermore, that neural patterns in both rEC and raHC are simultaneously associated with judgments of information flow, suggests people draw on maps encoded in the rEC and raHC to infer the dynamics of social transmission within the network.

If the rEC and raHC support tracking of information flow primarily by encoding efficient maps of the social network for subsequent retrieval, then we might expect that failures to encode these maps render the neural codes in these regions uninformative for navigational inference. To test this, we leverage the considerable individual differences in how strongly participants encode a Katz representation in rEC and raHC, and quantify them by extracting participant-specific beta coefficients for Katz from the random slopes of mixed-effects models predicting neural pattern similarity in these regions. We then ask whether inferences in the Information Flow Task are better predicted by neural pattern similarity in people who encoded Katz more strongly in these regions. As predicted, we find that our ability to use neural pattern (dis) similarity to predict a participant's behavioral inferences about information flow depends on how strongly participants encode a Katz representation in our regions of interest [modulation of rEC patterns: β = 0.172, 95% CI = (0.008, 0.336), z = 2.058, P = 0.040; modulation of raHC patterns: β = 0.177, 95% CI = (0.089, 0.266), z = 3.938, P < 0.001; Fig. 3 *D* and *E*].

For someone whose rEC neural patterns strongly differentiate network members based on Katz communicability (i.e., more negative Katz encoding), these neural codes reliably predict their inferences about whether information will spread from one member to another [simple effect β = −0.455, 95% CI = (−0.740, −0.171), z = −3.136, P = 0.002 at Katz encoding = mean − 1SD]. For someone whose rEC neural patterns are unrelated to Katz and thus contain no information about connectivity (i.e., Katz encoding = 0), rEC pattern dissimilarity for a dyad is unrelated to participants' inferences about information flow [simple effect β = −0.064, 95% CI = (−0.268, 0.140), z = −0.612, P = 0.540].

Similarly, for someone whose raHC neural representations exhibit greater similarity for dyads with greater Katz communicability (i.e., more positive Katz encoding), these codes reliably predict their inferences about whether information will spread [simple effect β = 0.539, 95% CI = (0.320, 0.759), z = 4.817, P < 0.001 at Katz encoding = mean + 1SD]. For someone whose raHC are unrelated to Katz and thus contain no information about connectivity (i.e., Katz encoding = 0), raHC pattern similarity between members of the dyad is again unrelated to participants' inferences about information flow [simple effect β = 0.036, 95% CI = (−0.126, 0.197), z = 0.433,

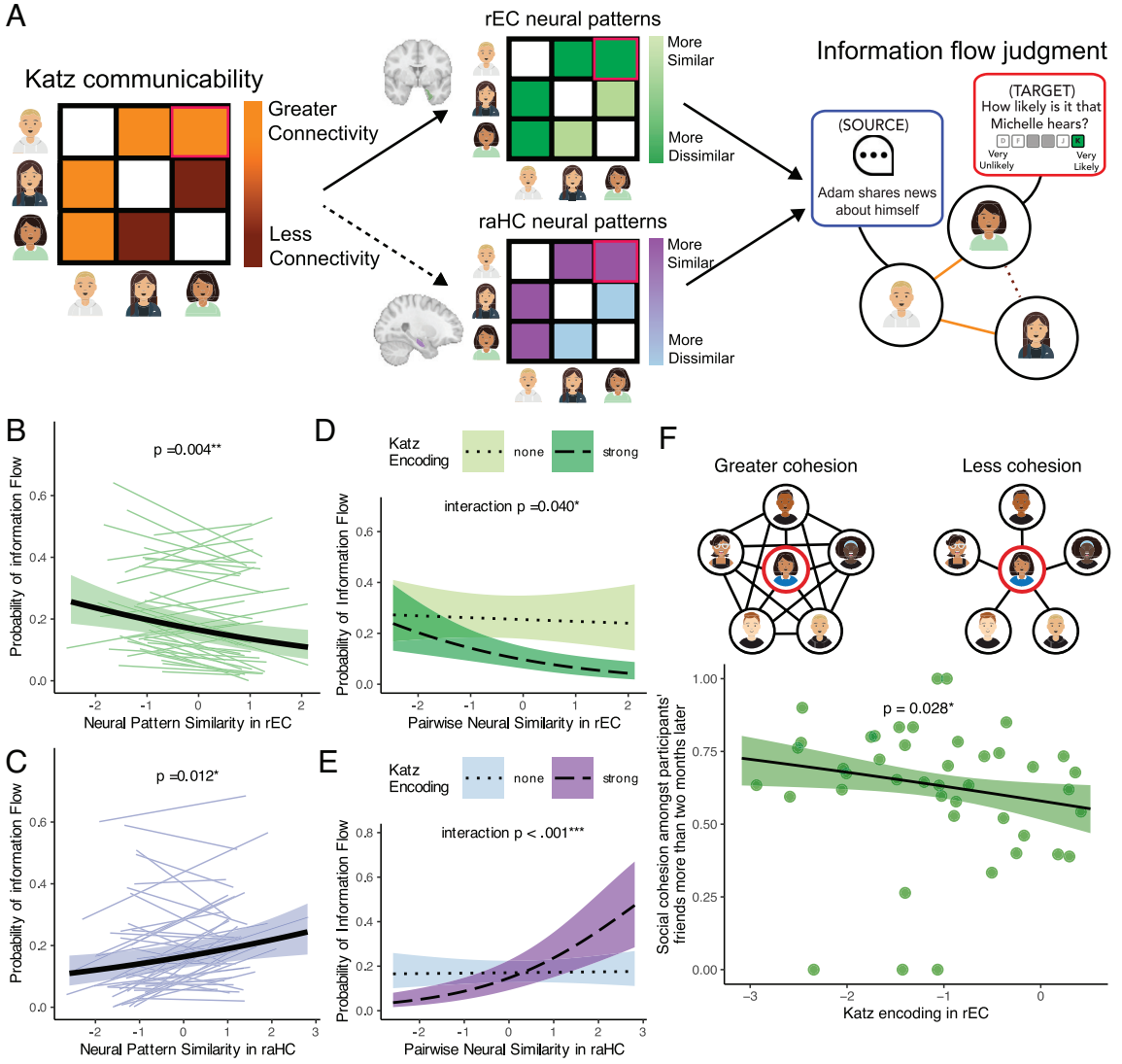

Fig. 3. Neural representation of abstract relational maps supports social navigation. (*A*) Predicted association between neural representations and information flow judgment. Because Katz communicability is encoded as greater pattern dissimilarity in rEC and greater similarity in raHC, inferences of information flow are supported by pattern dissimilarity in rEC and pattern similarity in raHC, we expect information flow judgments to be supported by pattern dissimilarity in rEC and pattern similarity in raHC. (*B*) Group-level effect in rEC. As Katz communicability is encoded using a dissimilarity code in rEC, inference of information flow is supported by pattern dissimilarity. (*C*) Group-level effect in raHC. As Katz may rely on a similarity code in raHC, inferring information flow is supported by pattern similarity. Each thin line in B-C indicates the best-fit line for each individual participant from a mixed-effects linear model fitted hierarchically to the data. Solid black lines and shaded regions in *B* and *C* indicate the mean and 95% CI of the average model estimated predictions respectively for the group. (*D*) Individual differences in rEC. The more strongly an individual's rEC encodes Katz communicability, the more strongly rEC predicts inference of information flow. (*E*) Individual differences in raHC. The more strongly an individual's raHC encodes Katz communicability, the more strongly raHC predicts inference of information flow. Lines and shaded regions in D–E indicate the mean and 95% CI of model-estimated probabilities of information flow between two network members, as a function of neural pattern similarity and level of Katz encoding in the respective regions. Dotted lines and light shading indicate the model-estimated probabilities when participants' neural patterns do *not* encode Katz (none: neural patterns ~ Katz $\beta = 0$). In contrast, dashed lines and dark shading indicate model estimates when participants exhibit strong levels of Katz encoding (strong: rEC patterns track very negatively with Katz, mean − 1SD: $\beta = -2.274$; raHC patterns track very positively with Katz, mean + 1SD: $\beta = 2.840$). (*F*) Abstract maps of multistep relations in rEC afford effective social brokerage. The degree to which an individual encodes Katz communicability in the rEC predicts greater social cohesion among their friends (i.e., more effective brokerage) more than two months later. Each point represents a single participant. Because rEC represents the multistep connectivity between dyads in the network using a dissimilarity code, negative values of Katz encoding (x-axis) indicate greater encoding strength. The black line and shaded region indicate the predicted mean and 95% CI from an ordered beta regression model. Having controlled for social cohesion among friends when neural activity was measured, this association can be effectively interpreted as the relative change in social cohesion over time. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. Icons: getavataaars.com

$P = 0.665$]. Cross-validated model comparisons and additional analyses further illustrate the robustness of these findings (*SI Appendix*, Note S5). Together, these results provide strong evidence that the rEC and raHC's encoding of multistep connections within the social network is consequential for tracking information flow.

rEC Maps of the Social Network Enable Effective Social Brokerage. There are, of course, other types of social navigation problems that go beyond inferences about information flow. Another notable feature of social networks is that each member and their social ties actively contribute to its evolution over time. Thus, we might also ask whether abstract neural maps of multistep relations afford individuals a unique ability to shape the social communities they are a part of. As we have already illustrated, possessing an abstract neural map enables the identification of multistep connectivity between people. Naturally, this also includes the connectivity among one's friends. With the spontaneous availability of this knowledge in their daily lives, one might then be able to subsequently broker

the formation of friendships between friends who are currently unconnected to each other to foster a more cohesive social community, which ultimately affords the individual greater social capital (1). To test this, we recontacted 176 participants from our initial network (N = 187) in the following semester at college, two months after neuroimaging, to identify their friendships again. This allows us to investigate whether the encoding of abstract neural maps earlier in the year is associated with greater social cohesion among participants' friends more than two months later.

Within a social network, the cohesion among an individual's friends can be quantified by (the inverse of) the average number of connections separating each pair of their friends if the participant were to be removed from the network—a metric known as local efficiency (56, 57). In other words, greater distances between a participant's friends if the participant is removed from the network indicate poor cohesion among the group. In contrast, a maximally cohesive social group would comprise individuals who are all mutually friends with each other. Confirming our hypothesis, we find that participants who more strongly encode an abstract neural map in the rEC (i.e., more negative Katz encoding) have more cohesive friend groups later in the academic year [β =–0.212, 95% CI = (–0.402, –0.023), z = –2.197, P = 0.028; Fig. 3*F* and *SI Appendix*, Table S10]. This was not the case for raHC encoding [β = 0.102, 95% CI = (–0.020, 0.223), z = 1.644, P = 0.100], or the control regions of V1 (Ps > 0.330, *SI Appendix*, Table S11). Having controlled for the social cohesion of their friend group prior to the neuroimaging session in these analyses, our findings imply that the availability of the rEC neural maps facilitates *increases* in the cohesion of one's social group: people who strongly encode these neural maps of their social network possess more cohesive friend groups independent of initial levels of cohesion. Moreover, we find that these effects are robust across numerous supplementary analyses (*SI Appendix*, Tables S12 and S13 and Note S6). The spontaneous availability of an abstract neural map of people's network in the rEC seems to afford people the ability to draw on the kinds of relational information necessary to broker the formation of friendships between their friends, fostering greater social cohesion in their immediate communities over time.

Discussion

People are remarkably proficient at navigating a wide array of social situations that require them to keep track of the intricate web of relations between members of a social network (21, 22, 25, 58). However, it has remained unknown exactly how the human brain organizes such forms of relational information into a useful representation for flexible behavior in the real-world, especially when networks are large and complex and direct observations about specific relationships are scarce. We combined computational modeling, functional neuroimaging, and longitudinal social network analysis of a large and emerging real-world network to answer this question in a naturalistic and ecologically valid setting. Despite constraints on being able to directly acquire knowledge about all other friendships in one's community (21, 23), we find that people correctly identify friendships in their personal social network far above chance levels. Consistent with recent work (21, 22), people accomplish this feat by constructing abstract representations of their social networks which integrate not only direct knowledge of friendships but also predictive inferences about unobserved (but likely) relations between people.

Our results also reveal the precise format of these representations. Prior work using laboratory tasks has reported evidence for people employing the Successor Representation in nonsocial (6, 20, 38, 39) and in social domains (21, 22). However, in this latter work, the implicit assumptions of the SR (about how social relationships are encoded) have remained relatively unexamined. Chief among them is the assumption that people represent others' relationships by normalizing the strength of their connections across the number of ties they maintain, which in turn makes the improbable prediction that people discount the strength of others' relationships proportionally based on their popularity (i.e., popular people do not have strong relationships). Indeed, we find that the Katz model (25, 32), which prioritizes the independent—rather than relative—encoding of people's concurrent relationships (30, 31, 33), handily outperforms the Successor Representation in explaining our participants' behavior. That Katz is a more likely format for representing knowledge about social relationships within a network where people can be concurrently connected to many others suggests that representations of real-world social environments are naturally structured to capture the specific dynamics of everyday social phenomena, such as how information simultaneously flows across multiple chains of relationships.

Furthermore, we provide direct evidence that a Katz-like abstract map of the social network is encoded in regions of the medial temporal lobe—primarily, the right entorhinal cortex (rEC). We also find suggestive evidence that the right anterior hippocampus (raHC) might similarly track a Katz-like cognitive map of the social network, though we emphasize that these effects are much less robust and therefore interpret them cautiously. Extant work has established that these same regions are implicated in constructing and maintaining map-like representations of relational information (3, 5–18, 29), including representations that resemble the SR in other domains (6, 8). Our finding that real-world social network structure is encoded in the MTL adds to this literature and argues for domain-general mapping functions of this region. However, it also raises the question of whether the format of representation (Katz vs SR) depends on specific properties of the relational structures (social vs nonsocial) they encode. In fact, some existing work implicating MTL in encoding the SR in nonsocial domains, relied on a modified version of the SR that more closely resembles Katz communicability (59). This hints at the possibility that Katz might be a more domain-general format for relational representation, though future work would have to more carefully and rigorously test this hypothesis across different domains. Regardless, our finding that people's neural representations of their social networks encode Katz-like multistep connectivity illuminates how common neural mechanisms of abstraction might underpin flexible and adaptive inferences across different environments that range from simple grid-worlds to complex non-Euclidean social networks.

Moreover, we demonstrate the versatility of these neural representations for adaptive social behavior. We find that these abstract neural codes support people's inferences about how information spreads across their social network (assessed in a completely separate and independent task). This ability to track the flow of information through social ties within a vast network provides a potential mechanistic explanation for how people evaluate the reputational effects of their social behavior and flexibly adapt their actions to actively manage these considerations that govern social life (25, 34, 37, 60–63). Simultaneously, we show that these neural codes appear to afford individuals the ability to meaningfully shape the evolution of their immediate social communities. By tracking the multistep connectivity between one's own friends, these maps identify opportunities for people to broker connections between unconnected individuals within the group and foster greater social cohesion. While our work cannot directly show that people deploy these maps to bring their

communities together, we find some evidence that the spontaneous availability of these neural maps in a passive viewing task predicts future increases in social cohesion of one's social group. As a result, these effects do not simply indicate that people who infer strong connectivity between peers who share multistep relations (e.g., my friends who are connected through me) also accurately anticipate that these individuals are likely to become friends. Instead, it is the degree to which these inferences are spontaneously activated during people's interactions with others that predicts increases in the cohesion of their social community. Taking the preliminary evidence from exploratory social network analyses reported in our supplements, we cautiously speculate that this might be because spontaneous inferences about multistep connectivity between one's friends might promote actions to bring them into contact with one another more often, thereby brokering a direct friendship between those who are currently unconnected. However, the nature of our data precludes us from making strong causal inferences about these mechanistic processes. Given the sociopsychological benefits of inhabiting a highly cohesive social community for well-being (64, 65), our results lay the foundation for future work to more precisely explore and verify how possessing these neural maps might uniquely endow individuals with a powerful form of social capital (1, 66).

Altogether, these findings advance our collective understanding of how exactly the human brain supports our ability to skillfully navigate complex real-world social networks by simultaneously synthesizing and extending three distinct areas of research. Prior work has separately documented that 1) the brain spontaneously represents relational properties of social network members (48), 2) people deploy multistep abstraction to navigate social networks (21, 22, 25), and 3) the medial temporal lobe constructs and maintains abstract representations of relational features in simple grid-worlds (3, 5–18, 29). Our work demonstrates that a Katz-based model of multistep abstraction generates mental representations of a real-world social network which both accurately capture people's behavioral inferences during social navigation and predict their neural representations of peers in the MTL. Not only does this provide a more precise characterization of how representations of social networks are formatted in the brain, but it also bolsters growing evidence for the domain-general role MTL plays in relational processing. In further documenting that possessing these neural representations confers potential social advantages, we additionally illustrate the adaptive utility of encoding abstract representations of our social environments (24, 26).

Yet, despite these insights, much remains to be understood about the precise neural mechanisms underpinning this capacity. Although we find evidence that the rEC encodes a map of people's social networks, it is unclear why it might deploy a dissimilarity code to do so. Pattern separation and differentiation primarily facilitate the resolution of interference between similar stimuli (52–55, 67), but have not yet been directly linked to the formation and maintenance of cognitive maps. Moreover, while the contributions of EC to pattern separation and differentiation in the HC are well documented (68–70), only recently has evidence emerged that the EC itself engages in pattern separation and differentiation (67). We speculate that the deployment of this dissimilarity code in the EC may reflect the utility of preventing interference between representations of distinct individuals who occupy similar structural positions within the social network—i.e., those who share a great deal of mutual and multistep connections (71, 72). Indeed, pairs of network members who exhibit strong multistep connectivity can also be considered relationally similar—i.e., occupying structurally similar social positions within the social network. Pattern differentiation based on multistep connectivity might thus afford individuals the ability to maintain more granular distinctions between their peers who run in the same social circles. Some recent work in the nonsocial domain aligns with this, reporting evidence that EC maintains more differentiated representations of items within a group compared to those across groups (73). Future work should more thoroughly interrogate this hypothesis.

Another potential avenue for future work concerns the preliminary evidence that neural representations of network relations in the rEC and raHC may be attuned to distinct components of network knowledge: The rEC seems to preferentially encode an inferential map of multistep connections while the raHC seems to preferentially encodes a more veridical map of relationships based on directly acquired information—though we emphasize that this latter claim is qualified by weaker evidence and warrants future replication. A potential source of uncertainty around the raHC representation comes from the fact that we have limited insight into both what (and how much) direct information people are able to acquire about others' social relationships "in the wild." This requires that we estimate what people might be observing based on where they sit in the network (21, 41). In reality, the relational information directly available to people is likely more diverse and idiosyncratic, and future work would have to develop more sophisticated techniques to circumvent the challenges of measuring these observations in real-world networks to confirm whether the brain separately encodes veridical memory for observed relationships.

Setting this aside, our work touches on past theories that medial temporal lobe representations might exhibit a potential gradient of abstraction, with more abstract representations encoded in the EC (3, 5) compared to the HC (6–8, 29). Such forms of complementary neural encoding might afford greater cognitive flexibility across a range of social decisions (4). Selective use of more abstract representations in EC can be prioritized when generalizing about unobserved or unknown relationships, and more veridical HC representations can be relied upon when greater precision is required. Future work investigating how people incrementally construct and update these integrated neural network representations from discrete, piecemeal information about pairwise relationships, and how these representations in the EC and HC are differentially reinstantiated during choice will help shed light on these open questions.

Materials and Methods

Overview. We recruited 198 first-year undergraduate students across three dormitories within the first month of their arrival on campus (N = 196 after 2 exclusions for not meeting eligibility; collected Sep 2022). At this initial time point, they completed demographic questionnaires, submitted photographs of themselves to be used as stimuli for subsequent data collection, and provided informed consent to be contacted for subsequent waves of data collection. Participants were then recontacted six times throughout the academic year to complete a Friendship Identification Task in which they identified their friends among the 195 other participants. They were also recontacted another three times to complete the Network Knowledge Task and Information Flow Task, and once to participate in a functional neuroimaging session.

Given our goal of identifying and characterizing the neural representations of the social network, we restricted our analyses to data collected during sessions directly adjacent to the functional neuroimaging session (N = 43 after 4 exclusions for data quality; collected Nov 2022). This included the second Friendship Identification Task collected approximately one month after the study began (N = 187; collected Oct 2022, T1), and the first sessions of the Network Knowledge Task and Information Flow Task (N = 100; collected Nov 2022 on separate occasions at least one week apart). Detailed description of the focal network and the tasks can be found in *SI Appendix* Note S1 and Table S14.

Finally, to assess how participants' social communities evolved over time, we examined changes in people's friendships between the second Friendship

Identification Task and the fourth friendship Friendship Identification Task (N = 176; collected Feb 2023, T2). To ensure that our observed associations were indicative of predictions and did not simply reflect changes in network structure that occurred between the second timepoint and our functional neuroimaging session, supplementary analyses confirm the robustness of these effects using the third Friendship Identification Task as an alternative baseline for comparison instead of the second (third Friendship Identification Task; N = 180; collected Dec 2022, T1.5). All procedures were approved by Brown University's Institutional Review Board, and informed consent was obtained from all participants. For 17-y-old minors, informed consent was obtained from their legal guardians, and assent was obtained from the participants.

Functional Neuroimaging. Results included in this manuscript come from preprocessing performed using *fMRIPrep* 21.0.0 (RRID:SCR_016216) (74, 75), which is based on *Nipype* 1.6.1 (RRID:SCR_002502) (76, 77) (see *SI Appendix*, Note S1 for full details). We performed Representational Similarity Analysis (RSA) (78) on our regions of interest (ROIs) with the Python implementation of RSA Toolbox (79). To obtain unbiased and reliable estimates of neural similarity, we computed cross-validated Mahalanobis (i.e., "crossnobis") distances (51). Given that cross-validating estimates of neural pattern similarity across runs produces an unbiased measure, the crossnobis estimator tends to be more reliable than other estimators through its use of multivariate noise normalization, which leverages the first-level GLM residuals to account for noise covariance. The "raw" crossnobis estimates reflect dissimilarity, so in our statistical analyses, we applied a sign-flip such that more positive (less negative) values now reflect greater similarity for convenience.

Computational Modeling.

Models of representation. We considered four candidate models of representation: Perfect Observation, Limited Observation, Katz Communicability (Katz), and the Successor Representation (SR). The common assumption across these models is that people can acquire information about others' relationships in the network. Using participants' responses in the Friendship Identification Task, we constructed a matrix, M, of size N × N, where N corresponds to the number of network members (N = 187) and the value in a cell of this matrix, $M(i,j)$, indicates whether members i nominated member j as their friend (0: not friend; 1:friend). Because not all friendship nominations are reciprocated, we only use mutually identified friendships to construct the ground-truth structure of the social network, represented by the adjacency matrix, A. The Perfect Observation model then simply assumes that participants' mental representations of the network reflect this ground-truth structure, A (Eq. **1**).

$$A(i,j) = \begin{cases} 1, & M(i,j) = 1 \wedge M(j,i) = 1 \\ 0, & otherwise \end{cases}. \quad [1]$$

In contrast, the Limited Observation model instead assumes that people are constrained in their ability to acquire information about all the mutual friendships in the network. Here, we note two constraints. First, we expect that people are most likely to observe social interactions occurring within their immediate "social circle" and are less and less likely to observe more distal relations (41), and estimate a parameter, $\omega \in [0,1]$, to quantify how strongly observations are affected by network distance, d. When $\omega \to 0$, participants only have information about their own friendships, whereas when $\omega \to 1$, they possess greater knowledge about relationship between dyads increasingly further away from them. Second, we consider the possibility that these observations do not depend on the strict reciprocity of relations, both between the observer and the dyad observed, and between members of the dyad themselves. Specifically, we consider that a person–e.g., Frank–is likely to observe someone they identify as their friend–e.g., Jack, to gain insight into who Jack identifies as their friend, even if Jack does not reciprocate Frank's friendship. Through this, Frank possibly finds out that Jack is friends with Max, and to a lesser extent, also who Max identifies as their friend.

Together, this produces a systematically biased representation, A', that is largely dependent on the outbound edges of the directed social network, M, where observations of the relationship between network members i and j, depends first and foremost on if either of them identifies the other as a friend (Eq. **2**). If neither do, the model assumes that people have no evidence of a relationship between the dyad. If only one member of the dyad identifies the other as their friend, the likelihood of the observation depends on the minimum number of outbound edges between the participant and the member who nominated the other, d. If both members of the dyad mutually identify the other as their friend, the likelihood of the observation depends on the minimum number of outbound edges that separate the participant from the closer member of the dyad, $min(d_i, d_j)$, Additional model comparisons suggest that this assumption that people's observations of the network depend on outbound edges better captures participants' Network Knowledge than models based on mutual edges and inbound edges (*SI Appendix*, Note S7).

$$A'(i,j) = \begin{cases} \omega^{d_i}, & M(i,j) = 1 \wedge d_i \le d_j \\ \omega^{d_j}, & M(j,i) = 1 \wedge d_i > d_j \\ 0, & otherwise \end{cases}. \quad [2]$$

Our models of abstraction–i.e., Katz and SR–similarly assume that people are limited in their ability to directly acquire information about all the mutual friendships like the Limited Observation model, but additionally abstract over these observations to generate inferences based on the multistep connections between each pair of network members. The Katz model abstracts directly over the biased adjacency matrix, A', to produce the Katz communicability matrix, KC (Eq. **3**) (25, 30–33).

$$KC = \sum_{k=0}^{\infty} \alpha^k A'^k. \quad [3]$$

In plain terms, when given the ground-truth adjacency matrix, A, we can compute the powers of this matrix A^k which produces matrices that indexes the number of k-step relations separating each pair of network members (or alternatively, the number of k-step paths connecting the pair). When applied to the biased adjacency matrix, A', A'^k analogously reflects the extent of k-step connectivity between two pairs, qualified by the participants' limited ability to observe each step. The Katz communicability is thus the weighted sum of all the multistep connections between each pair of network members, where each multistep connection is discounted by α based on the number of steps it comprises. Because α has a theoretical upper bound equal to the inverse of the maximum eigenvalue of the matrix, A', we can reexpress the Katz communicability as Eq. **4**:

$$KC = \sum_{k=0}^{\infty} \left(\gamma \cdot \frac{1}{\lambda}\right)^k A'^k, \quad [4]$$

where λ is the maximum eigenvalue of A' and $\gamma \in [0,1)$ such that γ indexes the degree to which a participant integrates (i.e., abstracts over) an increasing number of steps when representing pairwise associations in the network. When $\gamma \to 0$, the representation only contains direct observations of pairwise relations consistent with memorization. As $\gamma \to 1$, the representation becomes increasingly abstract, reflecting longer-range connections between a pair of individuals.

The SR, much like the Katz, abstracts over the observable pairwise relations in the network (8, 20, 38, 39). However, instead of abstracting directly over the biased adjacency matrix, A', it first normalizes it into a transition matrix, T' (Eq. **5**), which represents the relative probability that member i is friends with member j, as opposed to all other network members in a network of size N.

$$T'(i,j) = \frac{A'(i,j)}{\sum_{k=i}^{N} A'(i,k)}. \quad [5]$$

The SR is then computed as the weighted sum of the multistep transition probabilities between each pair of network members, where each multistep transition is discounted by $\gamma \in [0,1)$ based on the number of steps it comprises (Eq. **6**).

$$SR = \sum_{k=0}^{\infty} \gamma^k T'^k. \quad [6]$$

Additional information about model-fitting, evaluation, and comparison can be found in *SI Appendix*, Note S1.

Data, Materials, and Software Availability. All necessary data and code necessary to reproduce the results are openly accessible at https://osf.io/adm8v/ (80). All data provided are deidentified and labeled with randomly assigned pseudonyms to facilitate matching across datasets.

ACKNOWLEDGMENTS. We would like to thank Isabella Aslarus, Yi-Fei Jerry Hu, Elizabeth Duchan, Maya Mazumder, Kayleigh Danowski, Jonathan Palfy, Vera Poyraz, Samantha Shulman, Sofia Vaca Narvaja, and Jenny Wang for their indispensable assistance in data collection. Part of this research was conducted using computational resources and services at the Center for Computation and Visualization, Brown University. Advanced access to these computing resources was supported by NIH award 1S10OD025181. This work is supported by the NSF award 2123469 to O.F.H. and A.B.

Author affiliations: [a]Department of Cognitive and Psychological Sciences, Brown University, Providence, RI 02912; and [b]Robert J. and Nancy D. Carney Institute for Brain Science, Brown University, Providence, RI 02912

1. P. Bourdieu, "The forms of capital" in *Handbook of Theory and Research for the Sociology of Education*, J. Richardson, Ed. (Greenwood, 1986), pp. 241–258.
2. M. Peer, M. Hayman, B. Tamir, S. Arzy, Brain coding of social network structure. *J. Neurosci.* **41**, 4897–4909 (2021).
3. A. B. Baram, T. H. Muller, H. Nili, M. M. Garvert, T. E. J. Behrens, Entorhinal and ventromedial prefrontal cortices abstract and generalize the structure of reinforcement learning problems. *Neuron* **109**, 713–723.e7 (2021).
4. T. E. Behrens *et al.*, What is a cognitive map? Organizing knowledge for flexible behavior. *Neuron* **100**, 490–509 (2018).
5. S. Mark *et al.*, Flexible neural representations of abstract structural knowledge in the human Entorhinal Cortex. *eLife* **13**, RP101134 (2024).
6. A. Pudhiyidath *et al.*, Representations of temporal community structure in hippocampus and precuneus predict inductive reasoning decisions. *J. Cogn. Neurosci.* **34**, 1736–1760 (2022).
7. V. Samborska, J. L. Butler, M. E. Walton, T. E. J. Behrens, T. Akam, Complementary task representations in hippocampus and prefrontal cortex for generalizing the structure of problems. *Nat. Neurosci.* **25**, 1314–1326 (2022).
8. K. L. Stachenfeld, M. M. Botvinick, S. J. Gershman, The hippocampus as a predictive map. *Nat. Neurosci.* **20**, 1643–1653 (2017).
9. D. Aronov, R. Nevers, D. W. Tank, Mapping of a non-spatial dimension by the hippocampal–entorhinal circuit. *Nature* **543**, 719–722 (2017).
10. I. K. Brunec *et al.*, Multiple scales of representation along the hippocampal anteroposterior axis in humans. *Curr. Biol.* **28**, 2129–2135.e6 (2018).
11. C. F. Doeller, C. Barry, N. Burgess, Evidence for grid cells in a human memory network. *Nature* **463**, 657–661 (2010).
12. H. Eichenbaum, N. J. Cohen, Can we reconcile the declarative memory and spatial navigation views on hippocampal function?. *Neuron* **83**, 764–770 (2014).
13. D. Eilam, Of mice and men: Building blocks in cognitive mapping. *Neurosci. Biobehav. Rev.* **47**, 393–409 (2014).
14. T. Hafting, M. Fyhn, S. Molden, M.-B. Moser, E. I. Moser, Microstructure of a spatial map in the entorhinal cortex. *Nature* **436**, 801–806 (2005).
15. J. O'Keefe, L. Nadel, The hippocampus as a cognitive map (Oxford University Press, 1978).
16. M. Peer, I. K. Brunec, N. S. Newcombe, R. A. Epstein, Structuring knowledge with cognitive maps and cognitive graphs. *Trends Cogn. Sci.* **25**, 37–54 (2021).
17. S. A. Park, D. S. Miller, E. D. Boorman, Inferences on a multidimensional social hierarchy use a grid-like code. *Nat. Neurosci.* **24**, 1292–1301 (2021).
18. R. M. Tavares *et al.*, A map for social navigation in the human brain. *Neuron* **87**, 231–243 (2015).
19. E. D. Boorman, S. C. Sweigart, S. A. Park, Cognitive maps and novel inferences: A flexibility hierarchy. *Curr. Opin. Behav. Sci.* **38**, 141–149 (2021).
20. S. J. Gershman, The successor representation: Its computational logic and neural substrates. *J. Neurosci.* **38**, 7193–7200 (2018).
21. J.-Y. Son, A. Bhandari, O. FeldmanHall, Abstract cognitive maps of social network structure aid adaptive inference. *Proc. Natl. Acad. Sci.* **120**, e2310801120 (2023).
22. J.-Y. Son, M.-L. Vives, A. Bhandari, O. FeldmanHall, Replay shapes abstract cognitive maps for efficient social navigation. *Nat Hum Behav* **8**, 2167 (2024), 10.1038/s41562-024-01990-w.
23. O. FeldmanHall, M. R. Nassar, The computational challenge of social learning. *Trends Cogn. Sci.* **25**, 1045–1057 (2021).
24. R. Basyouni, C. Parkinson, Mapping the social landscape: Tracking patterns of interpersonal relationships. *Trends Cogn. Sci.* **26**, 204–221 (2022).
25. A. Xia, Y. Y. Teoh, M. R. Nassar, A. Bhandari, O. FeldmanHall, Knowledge of information cascades through social networks facilitates strategic gossip. *Nat. Hum. Behav.* **9**, 2169–2182 (2025).
26. O. FeldmanHall, J.-Y. Son, A. Bhandari, Abstract cognitive maps for complex social systems. *Curr. Dir. Psychol. Sci.* **34**, 09637214251342742 (2025), 10.1177/09637214251342742.
27. L. C. Freeman, Filling in the blanks: A theory of cognitive categories and the structure of social affiliation. *Soc. Psychol. Q.* **55**, 118–127 (1992).
28. M. E. Brashears, E. Quintane, The microstructures of network recall: How social networks are encoded and represented in human memory. *Soc. Netw.* **41**, 113–126 (2015).
29. J. C. R. Whittington *et al.*, The tolman-eichenbaum machine: Unifying space and relational memory through generalization in the hippocampal formation. *Cell* **183**, 1249–1263.e23 (2020).
30. A. Banerjee, A. G. Chandrasekhar, E. Duflo, M. O. Jackson, The diffusion of microfinance. *Science* **341**, 1236498 (2013).
31. P. Grindrod, M. C. Parsons, D. J. Higham, E. Estrada, Communicability across evolving networks. *Phys. Rev. E* **83**, 046120 (2011).
32. L. Katz, A new status index derived from sociometric analysis. *Psychometrika* **18**, 39–43 (1953).
33. G. Zamora-López, M. Gilson, An integrative dynamical perspective for graph theory and the analysis of complex networks. *Chaos: Interdiscip. J. Nonlinear Sci.* **34**, 041501 (2024).
34. L. Ellwardt, C. Steglich, R. Wittek, The co-evolution of gossip and friendship in workplace social networks. *Soc. Networks* **34**, 623–633 (2012).
35. M. Feinberg, R. Willer, M. Schultz, Gossip and ostracism promote cooperation in groups. *Psychol. Sci.* **25**, 656–664 (2014).
36. T. J. Grosser, V. Lopez-Kidwell, G. Labianca, A social network analysis of positive and negative gossip in organizational life. *Group Organ. Manag.* **35**, 177–212 (2010).
37. R. D. Sommerfeld, H.-J. Krambeck, D. Semmann, M. Milinski, Gossip as an alternative for direct observation in games of indirect reciprocity. *Proc. Natl. Acad. Sci.* **104**, 17435–17440 (2007).
38. P. Dayan, Improving generalization for temporal difference learning: The successor representation. *Neural Comput.* **5**, 613–624 (1993).
39. I. Momennejad *et al.*, The successor representation in human reinforcement learning. *Nat. Hum. Behav.* **1**, 680–692 (2017).
40. S. P. Borgatti, A. Mehra, D. J. Brass, G. Labianca, Network analysis in the social sciences. *Science* **323**, 892–895 (2009).
41. E. Breza, A. G. Chandrasekhar, A. Tahbaz-Salehi, Seeing the forest for the trees? An investigation of network knowledge. NBER Working Paper 24359 (2018), 10.3386/w24359.
42. E.-J. Wagenmakers, S. Farrell, AIC model selection using Akaike weights. *Psychon. Bull. Rev.* **11**, 192–196 (2004).
43. A. Gelman, J. Hwang, A. Vehtari, Understanding predictive information criteria for Bayesian models. *Stat. Comput.* **24**, 997–1016 (2014).
44. A. Vehtari, A. Gelman, J. Gabry, Practical Bayesian model evaluation using leave-one-out cross-validation and WAIC. *Stat. Comput.* **27**, 1413–1432 (2017).
45. S. Watanabe, Asymptotic equivalence of Bayes cross validation and widely applicable information criterion in singular learning theory. *J. Mach. Learn. Res.* **11**, 3571–3594 (2010).
46. L. Rigoux, K. E. Stephan, K. J. Friston, J. Daunizeau, Bayesian model selection for group studies–revisited. *Neuroimage* **84**, 971–985 (2014).
47. S. A. Morelli, Y. C. Leong, R. W. Carlson, M. Kullar, J. Zaki, Neural detection of socially valued community members. *Proc. Natl. Acad. Sci.* **115**, 8149–8154 (2018).
48. C. Parkinson, A. M. Kleinbaum, T. Wheatley, Spontaneous neural encoding of social network position. *Nat. Hum. Behav.* **1**, 1–7 (2017).
49. N. Zerubavel, P. S. Bearman, J. Weber, K. N. Ochsner, Neural mechanisms tracking popularity in real-world social networks. *Proc. Natl. Acad. Sci.* **112**, 15072–15077 (2015).
50. J. Diedrichsen *et al.*, Comparing representational geometries using whitened unbiased-distance-matrix similarity. *NBDT* **5**, 1–31 (2021).
51. A. Walther *et al.*, Reliability of dissimilarity measures for multi-voxel pattern analysis. *Neuroimage* **137**, 188–200 (2016).
52. A. J. H. Chanales, A. Oza, S. E. Favila, B. A. Kuhl, Overlap among spatial memories triggers repulsion of hippocampal representations. *Curr. Biol.* **27**, 2307–2317.e5 (2017).
53. C. B. Kirwan, C. E. L. Stark, Overcoming interference: An fMRI investigation of pattern separation in the medial temporal lobe. *Learn. Mem.* **14**, 625–633 (2007).
54. J. L. McClelland, B. L. McNaughton, R. C. O'Reilly, Why there are complementary learning systems in the hippocampus and neocortex: Insights from the successes and failures of connectionist models of learning and memory. *Psychol. Rev.* **102**, 419–457 (1995).
55. M. A. Yassa, C. E. L. Stark, Pattern separation in the hippocampus. *Trends Neurosci.* **34**, 515–525 (2011).
56. V. Latora, M. Marchiori, Efficient Behavior of Small-World Networks. *Phys. Rev. Lett.* **87**, 198701 (2001).
57. I. Vragović, E. Louis, A. Díaz-Guilera, Efficiency of informational transfer in regular and complex networks. *Phys. Rev. E* **71**, 036122 (2005).
58. J.-Y. Son, A. Bhandari, O. FeldmanHall, Cognitive maps of social features enable flexible inference in social networks. *Proc. Natl. Acad. Sci. U.S.A.* **118**, 1–11 (2021).
59. M. M. Garvert, R. J. Dolan, T. E. Behrens, A map of abstract relational knowledge in the human hippocampal–entorhinal cortex. *eLife* **6**, e17086 (2017).
60. R. D. Sommerfeld, H.-J. Krambeck, M. Milinski, Multiple gossip statements and their effect on reputation and trustworthiness. *Proc. R. Soc. B: Biol. Sci.* 275, 2529–2536 (2008).
61. T. D. Dores Cruz *et al.*, Gossip and reputation in everyday life. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **376**, 20200301 (2021).
62. M. L. Robbins, A. Karan, Who gossips and how in everyday life?. *Soc. Psychol. Personal. Sci.* **11**, 185–195 (2020).
63. J. Wu, D. Balliet, P. A. M. Van Lange, Reputation, gossip, and human cooperation. *Soc. Personal. Psychol. Compass* **10**, 350–364 (2016).
64. I. Kawachi, L. Berkman, Social cohesion, social capital, and health. *Soc. Epidemiol.* **174**, 290–319 (2000).
65. M. Kingsbury, Z. Clayborne, I. Colman, J. B. Kirkbride, The protective effect of neighbourhood social cohesion on adolescent mental health following stressful life events. *Psychol. Med.* **50**, 1292–1299 (2020).
66. N. E. Friedkin, Social cohesion. *Annu. Rev. Sociol.* **30**, 409–425 (2004).
67. Y. Zheng, J. Antony, C. Ranganath, R. C. O'Reilly, Recurrent inhibitory dynamics in the entorhinal cortex support pattern separation. bioXriv [Preprint] (2024), 10.1101/2024.11.14.623535 (Accessed 8 December 2024).
68. T. Amer, L. Davachi, Extra-hippocampal contributions to pattern separation. *eLife* **12**, e82250 (2023).
69. S. J. Guzman *et al.*, How connectivity rules and synaptic properties shape the efficacy of pattern separation in the entorhinal cortex–dentate gyrus–CA3 network. *Nat. Comput. Sci.* **1**, 830–842 (2021).
70. Z. M. Reagh *et al.*, Functional imbalance of anterolateral entorhinal cortex and hippocampal dentate/CA3 underlies age-related object pattern separation deficits. *Neuron* **97**, 1187–1198.e4 (2018).
71. E. A. Leicht, P. Holme, M. E. J. Newman, Vertex similarity in networks. *Phys. Rev. E* **73**, 026120 (2006).
72. F. Lorrain, H. C. White, Structural equivalence of individuals in social networks. *J. Math. Sociol.* **1**, 49–80 (1971).
73. C. Fernandez, J. Jiang, S.-F. Wang, H. L. Choi, A. D. Wagner, Representational integration and differentiation in the human hippocampus following goal-directed navigation. *eLife* **12**, e80281 (2023).
74. O. Esteban *et al.*, fMRIPrep. *Software* (2018), 10.5281/zenodo.852659.
75. O. Esteban *et al.*, fMRIPrep: A robust preprocessing pipeline for functional MRI. *Nature Methods* **16**, 111–116 (2018), 10.1038/s41592-018-0235-4.
76. K. Gorgolewski *et al.*, Nipype: A flexible, lightweight and extensible neuroimaging data processing framework in Python. *Front. Neuroinform.* **5**, 13 (2011).
77. K. J. Gorgolewski *et al.*, Nipype. *Software* (2018), 10.5281/zenodo.596855.
78. N. Kriegeskorte, J. Diedrichsen, Peeling the onion of brain representations. *Annu. Rev. Neurosci.* **42**, 407–432 (2019).
79. H. Nili *et al.*, A Toolbox for Representational Similarity Analysis.. *PLOS Comput. Biol.* **10**, e1003553 (2014).
80. Y. Y. Teoh, J-. Y. Son, A. Xia, A. Bhandari, O. FeldmanHall, Medial temporal lobe encodes cognitive maps of real-world social networks. Open Science Framework. https://osf.io/adm8v/. Deposited 29 March 2026.