Category-Based Routing in Social Networks: Membership Dimension and the Small-World Phenomenon

David Eppstein, Michael T. Goodrich, Maarten Löffler, Darren Strash, and Lowell Trott

Dept. of Computer Science, University of California, Irvine, USA

Abstract. A classic experiment by Milgram shows that individuals can route messages along short paths in social networks, given only simple categorical information about recipients (such as "he is a prominent lawyer in Boston" or "she is a Freshman sociology major at Harvard"). That is, these networks have very short paths between pairs of nodes (the so-called *small-world phenomenon*); moreover, participants are able to route messages along these paths even though each person is only aware of a small part of the network topology. Some sociologists conjecture that participants in such scenarios use a *greedy routing* strategy in which they forward messages to acquaintances that have more categories in common with the recipient than they do, and similar strategies have recently been proposed for routing messages in dynamic ad-hoc networks of mobile devices. In this paper, we introduce a network property called *membership dimension*, which characterizes the cognitive load required to maintain relationships between participants and categories in a social network. We show that any connected network has a system of categories that will support greedy routing, but that these categories can be made to have small membership dimension if and only if the underlying network exhibits the small-world phenomenon.

Keywords: membership dimension; small-world; category routing; social network

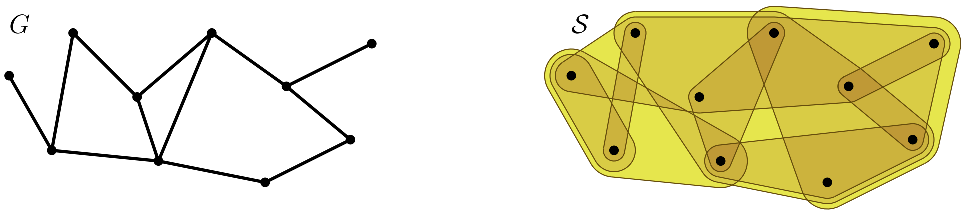
1 Introduction

In a pioneering experiment in the 1960's, Stanley Milgram and colleagues [16, 25, 30] empirically studied the ability of people in real-world social networks to route messages to their acquaintances, and used their studies to deduce properties of these networks. 296 randomly chosen individuals in Omaha, Nebraska and Wichita, Kansas were asked to forward a letter to a lawyer in Boston by using the following rule: send the letter to an acquaintance so that it progresses toward the recipient. Each acquaintance along the way is then told to forward the letter by this same rule. The results of these experiments reveal that, if a message gets to its recipient, it typically passes between at most six acquaintances[1]—and this observation has come to be called the *small-world phenomenon* [13, 31].

What is perhaps even more surprising than the existence of these short paths is the fact that human participants are able to efficiently route messages using only local information and simple facts about message targets, such as gender, ethnicity, occupation, name, and location.

As a way to study how humans can route such messages, several groups of sociology researchers have studied the importance of *categories*, that is, various groups to which people belong, in the small-world phenomenon. For instance, in the early 1970's, Hunter and Shotland [10] found that messages routed between participants who both belonged to the same category of people in a university (such as students, faculty, or administrators) had shorter paths than messages routed across such categories. Along these same lines, Killworth and Bernard [12] performed a set of experiments in the late 1970's they called *reverse small-world experiments*. In these experiments, they presented each participant with a list of messages for hundreds of targets, identified by the categories of town, occupation, ethnic background, and gender, and they asked each such participant to whom they would send each of these messages. One of the main conclusions of this study was that the choices people make in deciding on routes are overwhelmingly categorical in nature. In the late

[1] This observation has also led to the concept of "six degrees of separation" between all people on earth and the trivia game, "Six Degrees of Kevin Bacon," where players take turns trying to link performers to the actor Kevin Bacon via at most six movie collaborations.



(a) (b)

Fig. 1. A set of elements U (drawn arbitrarily as points in the plane). (a) The graph G on U. (b) The categories $\mathcal{S}$ on U. In this example, the membership dimension is 4, because no element is contained in more than 4 groups.

1980's, Bernard *et al.* [3] extended this work to identify which of twenty categories are the most important to people from various cultures for the sake of message routing. More recently, Watts *et al.* [32] present a hierarchical model for categorical organization in social networks for the sake of message routing. They propose groups as the leaves of rooted trees, with internal nodes defining groups-of-groups, and so on. They define an *ultrametric* on the vertices of each hierarchy (a distance function in which the distance between any two participants is determined by the level in the hierarchy of the smallest category containing both of them) and they conjecture that people use the minimum distance in one of their trees to make message routing decisions. That is, they argue that individuals can understand their "social distance" to a target as the minimum of the distances between them and the target in each of their hierarchical categories. Of course, such a determination requires some global knowledge about the structures of the various group hierarchies.

Although this previous work shows the importance of categories and of hierarchies of categories in explaining the small world phenomenon, it does not explain where the categories come from or what properties they need to have in order to allow greedy routing to work. Hence, this prior work leaves open the following questions:

- Which social networks support systems of categories that allow participants to route messages using the simple greedy rule of sending a message to an acquaintance who has more categories in common with the target?
- How complicated a system of categories is needed for this purpose, how much information about this system do individual participants need to know, and what properties of the underlying network can be used to characterize the complexity of the category system?

Our goal in this paper, therefore, is to address these questions by studying the existence of mathematical and algorithmic frameworks that demonstrate the feasibility of local, greedy, category-based routing in social networks.

1.1 Our Results

Inspired by the work of Watts *et al.* [32], we view a social network as an undirected graph $G = (U, E)$, whose vertices represent people and whose edges represent relationships, taken together with a collection, $\mathcal{S} \subset 2^U$, of categories defined on the vertices in G. Although the categories that we end up constructing in proving our results will have a natural hierarchical structure, we do not impose such a structure as part of our definitions. Figure 1 shows an example.

In addition, given such a social network, $G = (U, E)$ with a category system $\mathcal{S}$, we define the *membership dimension* of $\mathcal{S}$ to be

$$\max_{u \in U} |\{C \in \mathcal{S} \colon\ u \in C\}|,$$

that is, the maximum number of groups to which any one person in the network belongs. The membership dimension characterizes the *cognitive load* of performing routing tasks in the given system of categories—if the membership dimension is small, each actor in the network only needs to know a proportionately small amount of information about his or her own categories, his or her neighbors' categories, and the categories of each message's eventual destination. Thus, we would expect real-world social networks to have small membership dimension.

In this paper, we provide a constructive proof that a category system with low membership dimension can support greedy routing. Our results are not intended to model the actual formation of social categories, and we take no position on whether categories are formed from the network, the network is formed from categories, or both form together. Rather, our intention is to show the close relation between two natural parameters of a social network, its path length and its membership dimension. In particular:

- We show that the membership dimension of $(G, \mathcal{S})$ must be at least the diameter of G, $\mathrm{diam}(G)$, for a local, greedy, category-based routing strategy to work.
- Given any connected graph $G = (U, E)$, we show there is a collection, $\mathcal{S}$, of categories defined on the set, U, of vertices of G, such that local, greedy, category-based routing always works. Moreover, the membership dimension of $(G, \mathcal{S})$ in this case is $O((\mathrm{diam}(G) + \log |U|)^2)$.
- We show that some dependence between the category system and the underlying graph is essential, by proving that there does not exist a single category system that supports greedy routing regardless of its underlying graph.

Since the earliest work of Milgram [16, 25, 30], social scientists have generally believed the so-called "small world hypothesis" that the diameters of real-world social networks are bounded by small constants or by slowly growing functions of the network size. Under a weak form of this assumption, that the diameter is $O(\log |U|)$, our results provide a natural model for how participants in a social network could efficiently route messages using a local, greedy, category-based routing strategy while remembering an amount of information that is only polylogarithmic in the size of the network.

1.2 Previous Related Work

Greedy Routing. In addition to the greedy method described by Milgram [16, 25, 30] for routing in social networks, geometric greedy routing [8, 17] has been introduced in computer communications as a method to leverage the geographic location of nodes in ad-hoc and sensor networks in order to reduce the computational overhead of routing messages. In geometric greedy routing, vertices have coordinates in a geometric metric space. They use these coordinates to calculate the distances between the destinations of a message and their neighboring vertices; each message is routed greedily, to a neighbor that is closer to the message's destination. Not every geographic network has the property that this strategy will correctly route all messages to their destinations, so a number of techniques have been developed to assist such greedy routing schemes when they fail [4, 11, 18–20]. In a paradigm introduced by Rao et al. [29], virtual coordinates can also be introduced to overcome the shortcomings of real-world coordinates and allow simple greedy forwarding to function without the assistance of fallback algorithms. This approach has been explored by several other researchers [1, 14, 21, 28], who study various network properties that allow coordinates to be found that will cause greedy routing to succeed. In addition, several researchers also study the existence of *succinct* virtual coordinate systems [7, 9, 23, 26], where the number of bits needed to represent the coordinates of each vertex is polylogarithmic in the size of the network. If an assignment of virtual coordinates is succinct in this way, then the amount of computer memory needed to store the coordinates will be significantly smaller than the memory needed for a complete routing table that avoids the need for greedy routing. This notion of succinctness, and its motivation in reducing memory requirements, is closely analogous to our definition of the membership dimension for categorical greedy routing. Just as succinct bit representation is required to make greedy routing space-efficient, sociological routing requires low membership dimension to reduce the cognitive load on its participants and make it feasible for them to participate.

Almost all of this previous work on greedy routing in computer networks uses vertex coordinates in 2- and 3-dimensional Euclidean or hyperbolic spaces. However, one very recent exception to this restriction is work by Mei *et al.* [24], who study category-based greedy routing as a heuristic for performing routing in dynamic delay-tolerant networks of computing devices. Mei *et al.* assume that the network nodes have been organized into pre-defined categories based on their owners' interests. Their experiments suggest that using these categories for greedy routing is superior in practice to routing heuristics based on location or simple random choices. It is possible to interpret the categorical greedy routing techniques of Mei *et al.* and of this paper as being geometric routing schemes using virtual coordinates, where the coordinates of each node represent their category memberships. In this interpretation, the membership dimension of an embedding corresponds to the number of nonzero coordinates of each node, and our results show that such greedy routing schemes can be done succinctly in graphs with small diameter.

The Small-World Phenomenon Through an Algorithmic Lens. Like the work of this paper, Kleinberg [13] studies the small-world phenomenon from an algorithmic perspective. His approach takes an orthogonal direction from our work, however, in two ways. First, he focuses exclusively on location as the critical factor for supporting the small-world phenomenon (under a geometric metric), whereas our work focuses on greedy routing strategies based on categories and membership dimension. Second, his study takes vertex coordinates as a given and constructs the network from these coordinates based on geometry and random choices, whereas our approach takes the network as a given and studies the kinds of categorical structures needed to support category-based greedy routing.

In addition to this work by Kleinberg, many other researchers have proposed various different models for randomly generating graphs that possess properties similar to those in real-world social networks, such as being scale-free (obeying a power law in the degree distribution) or having small diameter. For instance, see [5, 6, 22, 27, 33].

2 Routing in Networks based on Categorical Information

In this section, we introduce a mathematical model of categorical greedy routing. This model defines in a precise way the routing strategy that we hypothesize people use to route messages in a real-world social networks, based on prior work [3, 10, 12, 32]. Additionally, we provide some basic definitions and properties that, when they hold for a network, allow us to guarantee the success of this routing strategy.

2.1 Basic definitions

Abstracting away the social context, let U be the universe of n people defining the potential sources, targets, and intermediates for message routes, and let $G = (U, E)$ be an undirected graph on U whose m edges represent pairs of people who can send messages to each other.

Definition 1 (diameter). *For any two elements $s, t \in U$, we define $\mathrm{sp}(s,t)$ to be the length of the shortest path between s and t in G. Then the diameter of G, denoted $\mathrm{diam}(G)$, is $\max_{s,t\in U} \mathrm{sp}(s,t)$, the maximum length of any shortest path in G. That is, it is the distance between the two vertices that are farthest from each other in G.*

In the greedy routing algorithms that we study, a central concept is a *neighborhood*, the set of participants that a message could be forwarded to.

Definition 2 (neighborhood). *For $s \in U$, we define the* neighborhood*, $N(s)$, to be the set of neighbors of s in G, that is,*

$$N(s) = \{u \in U \mid \{s, u\} \in E\}.$$

Moving from graphs to category systems, we define the membership dimension, a numerical measure of the complexity of a system of categories that is fundamental to our work.

Definition 3 (membership dimension). *Let $\mathcal{S} \subset 2^U$ be a set of subsets of U, which represent the abstract categories that elements of U can belong to. For a given $u \in U$, we define $\mathrm{cat}(u) \subset \mathcal{S}$ to be the set of groups to which u belongs:*

$$\mathrm{cat}(u) = \{C \in \mathcal{S} \mid u \in C\}.$$

Fig. 2. Illustration of the routing rule. v is a viable candidate for forwarding from u because v and w share more category memberships than u and w.

The membership dimension *of $\mathcal{S}$ is the maximum number of elements of $\mathcal{S}$ that any element of U is contained in, that is,*

$$\mathrm{memdim}(\mathcal{S}) = \max_{u \in U} |\,\mathrm{cat}(u)|.$$

As discussed in the introduction, there is reason to believe that in real world social networks and group structures $(G, \mathcal{S})$, both $\mathrm{diam}(G)$ and $\mathrm{memdim}(\mathcal{S})$ tend to be small.

2.2 The routing strategy

We now describe a simple category-based strategy to route a message from some node $s \in U$ to another node $t \in U$. The strategy is greedy, and therefore follows the *greedy routing rule*. We clarify the distance function following the definition:

Definition 4 (greedy routing rule). *If a node u receives a message M intended for a destination $w \neq u$, then u should forward M to a neighbor $v \in N(u)$ that is closer to w than u is, that is, for which $d(v, w) < d(u, w)$.*

As mentioned above, the distance function we study is category-based, and measures the number of shared groups of $\mathcal{S}$ that two nodes belong to. In particular, we define the distance $d(s, t)$ by the formula

$$d(s, t) = |\,\mathrm{cat}(t) \setminus \mathrm{cat}(s)|.$$

The backslash denotes the set-theoretic difference operator, so this distance function[2] measures the number of categories of the target that the current node does *not* share. This number decreases as the number of shared groups of $\mathcal{S}$ between the current node and the target increases. Figure 2 illustrates the routing rule. We refer to the greedy routing strategy that uses this distance function as ROUTING.

In real-world networks for which $\mathrm{memdim}(\mathcal{S})$ is small (as we conjecture), this strategy should be easy for participants to perform. A small $\mathrm{memdim}(\mathcal{S})$ makes it feasible for each participant to be aware of the categories to which he himself, his neighbors, and the target belong, and therefore allows the participants in the network to perform greedy routing with only a small cognitive load.

2.3 Successful routing

We now investigate under what conditions ROUTING can be successful in routing a message between any pair of nodes in a network. We identify several properties of a graph G and associated group structure $\mathcal{S}$ that directly influence the feasibility of the routing strategy.

For routing to be possible, G must be connected. But it seems natural to consider a stronger property, *internal connectivity*, which we define below.

Definition 5 (restriction). *If G is a graph, $\mathcal{S}$ is a category system for G, and C is a category in $\mathcal{S}$, then the* restriction *of G to C is the subgraph of G induced by C. That is, it is the graph with C as its vertex set and with an edge connecting every two vertices in C that are adjacent in G.*

[2] Note that d might not determine a metric space, because it need not necessarily be symmetric.

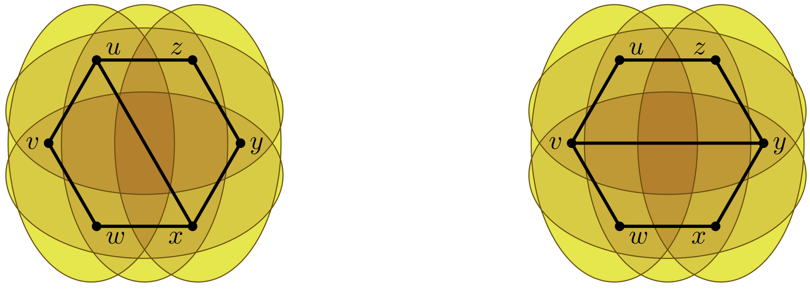


Fig. 3. Two examples with the same set of elements $U = \{u, v, w, x, y, z\}$ and categories $\mathcal{S} = \{\{u, v, w\}, \{x, y, z\}, \{u, w, x, z\}, \{u, v, y, z\}, \{v, w, x, y\}\}$. (a) An example that is internally connected, but not shattered: there is no neighbor of v that shares a region with y that v is not in. (b) An example that is shattered, but not internally connected: the induced graph of $\{u, w, x, z\}$ is not connected.

Definition 6 (internal connectivity). *A pair $(G, \mathcal{S})$ is* internally connected *if for each $C \in \mathcal{S}$, G restricted to C is connected.*

Figure 3(a) shows an example of an internally connected pair. This is a very natural property for sociological groups to exhibit. People belonging to the same group will have greater cohesiveness, and if a group fails the condition to be *internally connected*, then the group can be redefined sensibly to be the set of groups defined by their connected components.

Definition 7 (shattered). *A pair $(G, \mathcal{S})$ is* shattered *if, for all $s, t \in U$, $s \neq t$, there are a neighbor $u \in N(s)$ and a set $C \in \mathcal{S}$ such that C contains u and t, but not s.*

Figure 3(b) shows an example of a shattered pair. Note that in this definition, u and t could be the same node. This property falls out naturally from the instructions given in the real-world routing experiments of Milgram and others. In order for someone to advance a letter toward a target, there must be an acquaintance that shares additional interests with the target. Indeed, we now show that the shattered property is necessary for ROUTING to work.

Lemma 1. *If $(G, \mathcal{S})$ is not shattered, then ROUTING does not correctly route messages between all pairs of vertices.*

Proof. Suppose that $(G, \mathcal{S})$ is not shattered. That is, there exists a pair of vertices s and t, such that each category C that is shared by t and a neighbor of s is also shared by s. If this is the case, then it is not possible for any neighbor u of s to share strictly more categories with t as s does. Therefore, ROUTING will fail to route messages from s to t. □

Furthermore, if G is a tree, then these two properties of being shattered and of internal connectivity together are in fact *sufficient* for the routing strategy to always work.

Lemma 2. *If G is a tree, and $(G, \mathcal{S})$ is internally connected and shattered, then ROUTING is guaranteed to route messages correctly between every pair of vertices.*

Proof. Let s and t be any two vertices in G. Since G is a tree, there is one simple path from s to t. Let (u, v) be an edge on the path from s to t.

First, we claim that every category in $\mathcal{S}$ that contains both u and t also contains v. This follows from the assumption that $(G, \mathcal{S})$ is internally connected: any set $C \in \mathcal{S}$ with $u, t \in C$ must also contain v, since v is on the only path between u and t. Therefore, v is contained in at least as many sets in $\mathcal{S}$ with t as u is.

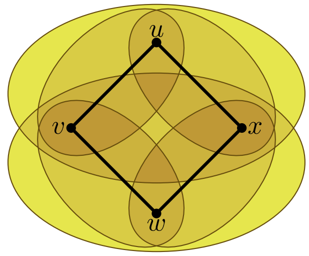


Fig. 4. The ROUTING strategy does not work in this graph, even though it is internally connected and shattered. This graph has just four vertices, $U = \{u, v, w, x\}$, connected in a cycle, taken together with the set of categories $\mathcal{S} = \{\{u, v, x\}, \{v, w, x\}, \{u, v\}, \{v, w\}, \{w, x\}, \{u, x\}\}$. However, ROUTING fails to route from v to x, since u is in 2 sets with x, v is in 2 sets with x, and w is in 2 sets with x.

However, by the assumption that $(G, \mathcal{S})$ is shattered, v must also share with t a category in $\mathcal{S}$ that does not contain u. Therefore v shares strictly more categories with t than u does, so ROUTING will correctly forward a message addressed to t from s to v.

Since we made no assumptions about s and t and showed that in each case ROUTING will always forward a message to the next vertex on a path to t, it follows that ROUTING succeeds for every pair of vertices. □

Although sufficient for routing in trees, the internally connected and shattered properties are not sufficient for ROUTING to work on arbitrary connected graphs. Figure 4 shows a counter-example—ROUTING is unable to route a message from the leftmost to the rightmost node, since there is no neighbor whose distance to the target is smaller.

3 Existence of Categories

In this section, we consider the following question: Is it possible to construct the family $\mathcal{S}$ so that ROUTING always works and $\mathcal{S}$ has low membership dimension?

We show that such a construction is always possible if we are given a connected graph as input. We also show that it is impossible to construct an $\mathcal{S}$ such that ROUTING will work if the graph is not known in advance.

3.1 Constructing $\mathcal{S}$ given G

Given a connected graph $G = (U, E)$ as input, we would like to construct a family $\mathcal{S} \subset 2^U$ so that ROUTING works, and the membership dimension of S is small. We concentrate foremost on constructions of category collections that are internally connected and shattered, because of the social significance of these properties. Nevertheless, even without these properties, we have the following lower bound.

Lemma 3. *Let G and $\mathcal{S}$ be a graph and a category system, respectively, such that ROUTING works for G and $\mathcal{S}$. Then* $\mathrm{memdim}(\mathcal{S}) \geq \mathrm{diam}(G)$.

Proof. Let s and t, be any two vertices of G, and let P be the path followed by ROUTING from s to t. An edge (u, v) can only be on P if $d(v, t) < d(u, t)$. Since $d(\cdot, \cdot)$ can only take integer values, $d(u, t) \geq d(v, t) + 1$. It follows by induction on the length of P that $d(s, t) \geq |P|$.

Now, by the definition of the diameter of a graph, there exists a pair of vertices $s, t \in U$ such that $sp(s, t) = \mathrm{diam}(G)$. Again, let P be the path that ROUTING follows from s to t; since the length of this path must be at least the length of a shortest path between the same two vertices, the length of P is at least $\mathrm{diam}(G)$.

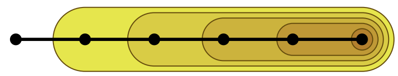


Fig. 5. The sets B_v for each vertex v in the path. The sets A_v are constructed symmetrically.

By definition, $d(s,t) = |\operatorname{cat}(t) \setminus \operatorname{cat}(s)|$, and $\operatorname{memdim}(\mathcal{S})$ is the maximum of $\operatorname{cat}(\cdot)$ over all elements. Putting these definitions together with the inequalities we have deduced between $d(s,t)$, $|P|$, and $\operatorname{diam}(G)$, we have

$$\operatorname{memdim}(\mathcal{S}) \geq |\operatorname{cat}(t)| \geq |\operatorname{cat}(t) \setminus \operatorname{cat}(s)| = d(s,t) \geq |P| \geq \operatorname{diam}(G),$$

as claimed. $\square$

For paths, this bound is tight:

Lemma 4. *If G is a path, then there exists a category system $\mathcal{S}$ for G such that $(G, \mathcal{S})$ is shattered and internally connected and such that* $\operatorname{memdim}(\mathcal{S}) = \operatorname{diam}(G)$.

Proof. Arbitrarily pick one of the two end vertices of G and let us refer to the vertices in G by their distance, 0 to $n-1$, from this vertex. For each vertex i, form two sets A_i and B_i, where $A_i = \{0, \ldots, i-1\}$ and $B_i = \{i+1, \ldots, n-1\}$, and let $\mathcal{S} = \bigcup_{v \in U} \{A_v, B_v\}$. Figure 5 illustrates this construction.

Each set in $\mathcal{S}$ consists of a path of vertices and therefore $\mathcal{S}$ is internally connected. $\mathcal{S}$ is also shattered, since for all s and t, s has a neighbor that shares either A_s or B_s with t, but s is not in these sets. To calculate $\operatorname{memdim}(\mathcal{S})$, note that each vertex i is contained in sets A_j for $0 \leq j < i$ and B_k for $k < i \leq n-1$. Therefore, each vertex is in exactly $n-1$ sets, which is $\operatorname{diam}(G)$. $\square$

A path is a special case of a tree. Therefore, whenever the given graph G is a path, it follows from Lemma 2 and Lemma 4 that it is possible to construct a category system $\mathcal{S}$ so that ROUTING works in G and so that the membership dimension $\operatorname{memdim}(\mathcal{S})$ equals $\operatorname{diam}(G)$,

There are also some other graphs, G, for which it is relatively easy to set up a category set, $\mathcal{S}$, that is shattered and internally connected in a way that supports the ROUTING algorithm. For example, in a tree of height 1 (i.e., a star graph), with root r, we could simply create a separate category containing the root r and each (leaf) child, plus a singleton category for each node. Every path in this tree clearly supports the ROUTING strategy. Note, however, that the membership dimension of this category system is high, since the root belongs to a linear number of categories. So even in this simple example, supporting the ROUTING strategy and achieving a small membership dimension is a challenge. Moreover, this challenge becomes even more difficult already for a tree of height 2, since navigating from any leaf, x, to another leaf, y, requires that the parent of x belong to more categories with y than x—and this must be true for *every* other leaf, y. Thus, it is perhaps somewhat surprising that we can construct a set of categories, $\mathcal{S}$, for an arbitrary binary tree that causes this network to be shattered and internally connected (so the ROUTING strategy works, by Lemma 2) and such that $\mathcal{S}$ has small membership dimension.

Lemma 5. *If G is a binary tree, then there exists a category system $\mathcal{S}$ such that $(G, \mathcal{S})$ is shattered and internally connected and such that* $\operatorname{memdim}(\mathcal{S}) = O(\operatorname{diam}^2(G))$.

Proof. We show how to construct $\mathcal{S}$ from G. Arbitrarily pick a vertex $r \in U$ of degree at most 2 and root the binary tree at r, so each vertex v has left and right children, $\operatorname{left}(v)$ and $\operatorname{right}(v)$, and let $\operatorname{height}(v)$ be the length of the longest simple path from v to any descendant of v. For each vertex v, we create a set S_v, containing v's descendants (which includes v). We further construct two families, L_v and R_v, using helper sets $L_{v,i}$ and $R_{v,i}$. Let $L_{v,i}$ (resp., R_{v_i}) consist of v, the vertices in v's left (right) subtree down to depth i, and all vertices in v's right (left) subtree. Then define

$$L_v = \{L_{v,i} \mid \operatorname{depth}(v) \leq i \leq \operatorname{depth}(v) + \operatorname{height}(\operatorname{left}(v))\}.$$

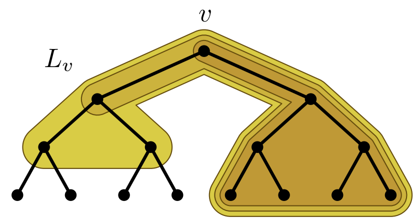


Fig. 6. Showing the collection of sets L_v for a small example subtree at v.

Figure 6 illustrates this. The family R_v is defined symmetrically. Our $\mathcal{S}$ is then defined as

$$\mathcal{S} = \bigcup_{v \in U} \{S_v\} \cup L_v \cup R_v.$$

By construction, each set in $\mathcal{S}$ is a connected subgraph of G and therefore $\mathcal{S}$ is internally connected. We can also see that $\mathcal{S}$ is shattered as follows. If s is an ancestor of t, then s's child u on the path to t is contained in set S_u which contains u, t, and not s. Otherwise, let v be the lowest common ancestor of s and t, and assume without loss of generality that s in v's left subtree; then s's parent is in $L_{v,\mathrm{depth}(s)-1}$ with t, and s is not.

We now analyze the membership dimension of this construction. Let v be a vertex, and let ancestors(v) be the set of v's ancestors. For $u \in \mathrm{ancestors}(v)$, $v \in S_u$, and v belongs to $O(\mathrm{height}(u))$ sets of L_u and R_u. Then v belongs to $O\left(\sum_{u \in \mathrm{ancestors}(v)} \mathrm{height}(u)\right)$ sets, which is $O(\mathrm{diam}^2(G))$ for any v. ◻

We now show how to extend this result to arbitrary trees. Our technique involves an application of *weight balanced binary trees* [2, 15].

Definition 8 (weight balanced binary tree). *A weight balanced binary tree is a binary tree that stores weighted items in its leaves. If item i has weight w_i, and all items have a combined weight of W then item i is stored at depth $O(\log(W/w_i))$.*

Lemma 6. *Let T be an n-node rooted tree with height h. We can embed T into a binary tree such that the ancestor–descendant relationship is preserved, and the resulting tree has height $O(h + \log n)$.*

Proof. Let n_u be the number of descendants of vertex u in T. For each vertex u in T that has more than two children, we expand the subtree consisting of u and u's children into a binary tree as follows. Construct a weight balanced binary tree B on the children of u, where the weight of a child v is n_v. We let u be the root of B. Each child v of u in the original tree is then a leaf at depth $\log(n_u/n_v)$ in B. Performing this construction for each vertex u in the tree expands T into a binary tree with the ancestor–descendant relationship preserved from T.

Furthermore, each path from root to leaf in T is only expanded by $\log(n)$ nodes, which we can see as follows. Each parent-to-child edge (u, v) in T is replaced by a path of length $O(\log(n_u/n_v))$. Therefore for each path P from root r to leaf l in T, our construction expands P by length $O(\sum_{(u,v)\in P} \log(n_u/n_v))$, which is a sum telescoping to $O(\log(n_r/n_l)) = O(\log n)$. Therefore, the height of the new binary tree is $O(h + \log n)$. ◻

Combining this lemma with Lemma 2, we get the following theorem.

Theorem 1. *Given a tree T, it is possible to construct a family $\mathcal{S}$ of subsets such that* ROUTING *works for T and* $\mathrm{memdim}(\mathcal{S}) = O((\mathrm{diam}(T) + \log n)^2)$.

Proof. Arbitrarily root T and embed T in a binary tree B using the method in Lemma 6. Then B has height $O(\mathrm{diam}(T) + \log n)$, and diameter $\mathrm{diam}(B) = O(\mathrm{diam}(T) + \log n)$. Applying the construction from Lemma 5

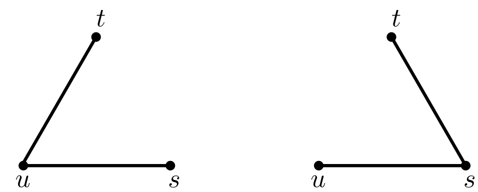


Fig. 7. Two connected graphs on the same vertex set. Given the underlying vertex set, we cannot form a set of groups so that our greedy strategy routes from s to t in the left graph and from u to t in the right graph.

to B gives us a family $\mathcal{S}_B$ with $\mathrm{memdim}(\mathcal{S}_B) = O((\mathrm{diam}(T) + \log n)^2)$. We then construct a family $\mathcal{S}_T$, by removing vertices that are in B but not T from the sets in $\mathcal{S}_B$. By construction, $(T, \mathcal{S}_T)$ is shattered and internally connected, and $\mathrm{memdim}(\mathcal{S}_T) \leq \mathrm{memdim}(\mathcal{S}_B) = O((\mathrm{diam}(T) + \log n)^2)$. By Lemma 2, ROUTING works on T with category sets from $\mathcal{S}_T$. □

We can further extend this theorem to arbitrary connected graphs, which is the main upper bound result of this paper.

Theorem 2. *If G is a connected graph, then there exists a category system $\mathcal{S}$ such that ROUTING correctly routes messages between all pairs of vertices and such that* $\mathrm{memdim}(\mathcal{S}) = O((\mathrm{diam}(G) + \log(n))^2)$.

Proof. Compute a low-diameter spanning tree T of G. This step can easily be done using breadth-first search, producing a tree with diameter at most $2\,\mathrm{diam}(G)$. We then use the construction from Theorem 1 on T. For greedy routing to work in a graph G, note that it is sufficient to show that it works in a spanning tree of G. Therefore, since ROUTING works in T, ROUTING also works in G. □

3.2 An Impossibility Result

It would be nice to construct a good group structure without knowing the structure of the graph in advance. Unfortunately, as we now show, this is impossible in general.

Theorem 3. *Given a set of vertices U, it is impossible to construct a set of groups $\mathcal{S}$ such that our greedy routing strategy works on all connected graphs with U as the vertex set.*

Proof. Consider the two graphs in Figure 7. For ROUTING to route from s to t in the left graph, u must share more groups with t than s does. However, to route from u to t in the right graph, s must share more groups with t than u does. Both of these events cannot happen simultaneously with one set of groups $\mathcal{S}$. Therefore ROUTING must fail in one of these two graphs. □

4 Conclusion and Open Problems

We have presented a construction of groups S on a connected graph G that allows a simple greedy routing algorithm, utilizing a notion of distance on group membership, to guarantee delivery between nodes in G. Such a construction will have membership dimension $O((\mathrm{diam}(G)+\log n)^2)$, which demonstrates a reasonably small cognitive load for the members of G.

There are several directions for future work. For example, while we have shown that the membership dimension must be minimally the diameter of G, it remains to be shown if the membership dimension must be the square of the diameter plus a logarithmic factor for arbitrary graphs. We conjecture that the square term is not strictly needed in the membership dimension in order for ROUTING to work. Our group construction is performed for a general graph by selecting a low diameter spanning tree and using the presented tree construction, so it may be possible that there is a group construction that has lower membership dimension and more efficient routing if it is constructed directly in G.

In addition, we observe that the groups that we construct in our upper-bound proofs have a natural nesting property that may correspond to a proximity-based way that people would organically form groups. It would be nice to verify or refute a hypothesis that people can organize themselves in such groups using local information and simple rules about how to form groups.

Finally, we took the perspective in this paper that all categories have equal weight with respect to routing tasks and that participants use a simple greedy routing algorithm based solely on increasing the number of categories in common with the target. One possible direction for future work would be to define and study a category-based routing strategy that allows participants to weight various categories higher than others, as in the work of Bernard *et al.* [3]. This could include giving higher consideration to smaller or more well connected groups. Another possible branch of further study might include analysis of the performance of this model when actors have only partial knowledge of the categories. A comparison could then be made between route lengths and level of category knowledge.

References

1. P. Angelini, F. Frati, and L. Grilli. An algorithm to construct greedy drawings of triangulations. In I. Tollis and M. Patrignani, editors, *Graph Drawing*, volume 5417 of *LNCS*, pages 26–37. Springer, 2009.
2. S. W. Bent, D. D. Sleator, and R. E. Tarjan. Biased search trees. *SIAM J. Computing*, 14(3):545–568, 1985.
3. H. R. Bernard, P. D. Killworth, M. J. Evans, C. McCarty, and G. A. Shelley. Studying social relations cross-culturally. *Ethnology*, 27(2):155–179, 1988.
4. P. Bose, P. Morin, I. Stojmenovic, and J. Urrutia. Routing with guaranteed delivery in ad hoc wireless networks. *Wireless Networks*, 7:609–616, 2001.
5. C. Cooper and A. Frieze. A general model of web graphs. *Random Structures & Algorithms*, 22(3):311–335, 2003.
6. P. Duchon, N. Hanusse, E. Lebhar, and N. Schabanel. Could any graph be turned into a small-world? *Theoretical Computer Science*, 355(1):96–103, 2006.
7. D. Eppstein and M. T. Goodrich. Succinct greedy graph drawing in the hyperbolic plane. In I. G. Tollis and M. Patrignani, editors, *Graph Drawing*, pages 14–25. Springer, 2009.
8. G. G. Finn. Routing and addressing problems in large metropolitan-scale internetworks. Technical report, ISI Research Report, 1987.
9. M. Goodrich and D. Strash. Succinct greedy geometric routing in the Euclidean plane. In Y. Dong, D.-Z. Du, and O. Ibarra, editors, *Algorithms and Computation*, volume 5878 of *LNCS*, pages 781–791. Springer, 2009.
10. J. E. Hunter and R. L. Shotland. Treating data collected by the "Small World" method as a Markov process. *Social Forces*, 52(3):321–332, 1974.
11. B. Karp and H. T. Kung. Gpsr: greedy perimeter stateless routing for wireless networks. In *6th ACM Cong. on Mobile Computing and Networking (MobiCom)*, pages 243–254, 2000.
12. P. Killworth and H. Bernard. Reverse small world experiment. *Social Networks*, 159(1), 1978.
13. J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proc. 32nd ACM Symp. on Theory of Computing (STOC)*, pages 163–170. ACM, 2000.
14. R. Kleinberg. Geographic routing using hyperbolic space. In *26th IEEE Conf on Computer Communications (INFOCOM)*, pages 1902–1909, May 2007.
15. D. E. Knuth. Optimum binary search trees. *Acta Informatica*, 1:14–25, 1971.
16. C. Korte and S. Milgram. Acquaintance networks between racial groups: Application of the small world method. *J. Personality and Social Psychology*, 15(2):101–108, 1970.
17. E. Kranakis, H. Singh, and J. Urrutia. Compass routing on geometric networks. In *11th Canadian Conf. on Computational Geometry*, pages 51–54, 1999.
18. F. Kuhn, R. Wattenhofer, Y. Zhang, and A. Zollinger. Geometric ad-hoc routing: of theory and practice. In *22nd ACM Symp. on Principles of Distributed Computing (PODC)*, pages 63–72, 2003.
19. F. Kuhn, R. Wattenhofer, and A. Zollinger. Asymptotically optimal geometric mobile ad-hoc routing. In *6th ACM Workshop on Discrete Algorithms and Methods for Mobile Computing and Communications (DIALM)*, pages 24–33, 2002.
20. F. Kuhn, R. Wattenhofer, and A. Zollinger. Worst-case optimal and average-case efficient geometric ad-hoc routing. In *4th ACM Symp. on Mobile ad hoc Networking & Computing (MobiHoc)*, pages 267–278, 2003.
21. T. Leighton and A. Moitra. Some results on greedy embeddings in metric spaces. *Discrete and Computational Geometry*, 44:686–705, 2010.

22. C. Martel and V. Nguyen. Analyzing Kleinberg's (and other) small-world models. In *23rd ACM Symp. on Principles of Distributed Computing (PODC)*, pages 179–188, 2004.
23. P. Maymounkov. Greedy embeddings, trees, and Euclidean vs. Lobachevsky geometry, 2006. `http://pdos.csail.mit.edu/~petar/papers/maymounkov-greedy-prelim.pdf`.
24. A. Mei, G. Morabito, P. Santi, and J. Stefa. Social-aware stateless forwarding in pocket switched networks. In *30th IEEE Conf on Computer Communications (INFOCOM)*, 2011.
25. S. Milgram. The small world problem. *Psychology Today*, 1(May):61–67, 1967.
26. R. B. Muhammad. A distributed geometric routing algorithm for ad hoc wireless networks. *Information Technology: New Generations, Third International Conference on*, 0:961–963, 2007.
27. M. E. J. Newman. Models of the small world. *Journal of Statistical Physics*, 101:819–841, 2000.
28. C. H. Papadimitriou and D. Ratajczak. On a conjecture related to geometric routing. *Theor. Comput. Sci.*, 344:3–14, November 2005.
29. A. Rao, S. Ratnasamy, C. Papadimitriou, S. Shenker, and I. Stoica. Geographic routing without location information. In *Proceedings of the 9th annual international conference on Mobile computing and networking*, MobiCom '03, pages 96–108, New York, NY, USA, 2003. ACM.
30. J. Travers and S. Milgram. An experimental study of the small world problem. *Sociometry*, 32(4):425–443, 1969.
31. D. J. Watts. Networks, dynamics, and the small-world phenomenon. *The American Journal of Sociology*, 105(2):493–527, 1999.
32. D. J. Watts, P. S. Dodds, and M. E. J. Newman. Identity and search in social networks. *Science*, 296:1302–1305, 2002.
33. D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393:440–442, 1998.