

On Leader Selection for Performance and Controllability in Multi-Agent Systems

Andrew Clark, Linda Bushnell, and Radha Poovendran

Department of Electrical Engineering, University of Washington, Seattle, WA, 98195, USA

Abstract—In a leader-follower multi-agent system (MAS), a set of leader agents act as external control inputs and are used to influence the dynamics of the remaining follower agents. Current approaches to selecting leaders are based on either achieving controllability of the follower agents or optimizing performance criteria such as robustness to noise, but not both. In this paper, we present a framework for selecting leaders based on joint consideration of controllability and performance. We first show that for the case where the number of nodes that can act as leaders is sufficient to guarantee controllability, the leader selection problem can be posed within a matroid optimization framework. For the case where the number of nodes that can serve as leaders is fixed and may not be sufficient for controllability, we introduce a new metric, the graph controllability index (GCI), defined as the fraction of network nodes that are controllable using the leader set. We prove that the GCI is a submodular function of the set of leader agents, leading to a submodular relaxation to the problem of achieving controllability. Our results are demonstrated using simulation study and compared to other leader selection algorithms, including random, average degree and descending order of degree based leader selection.

I. INTRODUCTION

Multi-agent systems (MAS) consist of a set of distributed nodes that exchange information and coordinate in order to achieve a shared task. In applications such as formation control [1], a MAS is controlled by a relatively small set of leader nodes whose state values (such as attitude, heading, and velocity) are determined by the system owner. The remaining nodes, known as followers, observe the states of their neighbors and update their states accordingly [2]. By updating their states based on the observed states of the leaders, the followers are influenced by the leader nodes. Through this process, the inputs from the leader nodes propagate through the system.

A key design parameter of leader-follower systems is the choice of leader agents, which impacts several aspects of the system performance. In [3], it was shown that the choice of leader agents determines whether the system is controllable. In [2], the choice of leaders was shown to affect the coherence of the system, defined as the ability of the system to achieve consensus in the presence of noise. The choice of leader nodes also influences the robustness of the system to link and agent failures [4]. Furthermore, even in the absence of noise and failures, the leader agents impact the rate at which the distributed control law converges [5].

Current design approaches to leader selection in MAS can be broadly classified into two categories. In the first category, leaders are chosen in order to optimize a continuous performance metric, such as the mean-square error due to

noise in the links between agents, within either a discrete [6] or continuous [7] optimization framework. While these approaches minimize the asymptotic error of the agent states, they do not guarantee that the selected agents can control the system.

The second category of leader selection methods focuses on achieving controllability using combinatorial algorithms. In [8], making use of results from [9] on the theory of structural controllability, a polynomial-time matching algorithm was introduced to select the minimum-size set of leader nodes necessary to control the system. While this algorithm guarantees controllability, it is not guaranteed to be optimal when even a single link experiences noise.

At present there is no design method for leader selection that incorporates both controllability and performance metrics. Moreover, when the number of nodes that can act as leaders is insufficient to control the entire network, there is currently no approach for selecting a leader set that controls as much of the network as possible. We aim to address these problems within a unifying framework.

We note that there are two types of leader selection design problems in MAS. In the first problem, a fixed number of nodes are chosen as leaders in order to achieve controllability and optimal performance. In the second problem, the number of leaders is not fixed and the minimum-size set of leader nodes must be selected to satisfy controllability and meet a given set of constraints on the performance metrics. In this paper, we introduce an analytical framework for addressing the first problem, leaving the second as future work.

Our contributions are two-fold. First, for the case where the number of nodes that can be chosen as leaders is sufficiently large to control the network, we formulate a broad class of leader selection problems within a matroid optimization framework in which controllability is a constraint. We then prove that controllability can be added as a constraint to several existing leader selection problems for different performance metrics, including leader selection to minimize noise and convergence error, with efficient algorithms that provide a provable bound on optimality.

Second, for the case where the number of leader nodes is insufficient to guarantee controllability, we introduce the *graph controllability index*, a submodular relaxation for the controllability criterion, and provide efficient, polynomial-time algorithms for selecting leaders in order to maximize controllability as well as performance. For the matroid optimization case, our results are illustrated via simulation study.

This paper is organized as follows. In Section II, we state the system model and provide background on matroid

theory. In Section III, we present our matroid optimization framework for leader selection. In Section IV, we introduce and derive a submodular relaxation for leader selection based on controllability. Section V contains our simulation study. Section VI concludes the paper.

II. BACKGROUND AND PRELIMINARIES

In this section, we define the system model and give background on graph matching and controllability. Relevant definitions of matroids and submodular functions are also reviewed.

A. System Model

A MAS consisting of n agents, indexed $V = \{1, \dots, n\}$, is considered. An arc (i, j) from agent i to agent j exists if agent j can receive control inputs from agent i . The set of nodes i such that (i, j) exists are the neighbors of j , denoted $N(j)$. For any subset $A \subseteq V$, $N(A) \triangleq \bigcup_{j \in A} N(j)$. The set of directed arcs is denoted E . We assume that the graph $G = (V, E)$ is strongly connected, implying that for every pair of nodes i and i' , there exists a directed path from i to i' . The graph topology G is assumed to be fixed over time.

Agent i maintains a time-varying state, denoted $x_i(t)$. A subset of leader agents, denoted $S \subseteq V$, receive their state values directly from the network owner. Each follower agent $i \in V \setminus S$ updates its state based on the current state values of its neighbors, $\dot{x}_i(t) = \sum_{j \in N(i)} W_{ij}x_j(t) + W_{ii}x_i(t)$. Letting $\mathbf{x}_F(t)$ denote the vector of follower node states and $\mathbf{x}_L(t)$ denote the vector of leader node states, the dynamics of the follower nodes can be written in vector form as

$$\dot{\mathbf{x}}_F(t) = W \begin{bmatrix} \mathbf{x}_F \\ \mathbf{x}_L \end{bmatrix} = A\mathbf{x}_F + B\mathbf{x}_L \quad (1)$$

where $W = [A \mid B]$.

B. Graph Matching

A bipartite graph is defined to be a graph $G = (V, E)$ such that $V = V_1 \cup V_2$, with $V_1 \cap V_2 = \emptyset$ and, for every $(i, j) \in E$, $i \in V_1$ and $j \in V_2$. In this case, we write $G = (V_1, V_2, E)$. A *matching* on a bipartite graph G is a mapping $m : V_1 \rightarrow V_2$ such that $m(v) \neq m(w)$ for $v \neq w$. The *cardinality* of the matching is defined to be $|m(V_1)|$. A matching is *maximal* if, after adding any edge to m , m is no longer a matching. A matching is *perfect* if $|m(V_1)| = |V_2|$. By definition, any perfect matching is maximal. The following theorem gives an equivalent condition for the existence of a perfect matching.

Theorem 1 (Hall Marriage Theorem [10]): Let $A \subseteq V_2$, and define $J(A) = \{i \in V_1 : (i, j) \in E \text{ for some } j \in A\}$. A perfect matching $m : V_1 \rightarrow V_2$ exists if and only if, for every $A \subseteq V_2$, $|A| \leq |J(A)|$.

C. Controllability

Definition 1 ([11]): The system defined by (1) is *controllable* if, for any initial state $\mathbf{x}_F(0) = \mathbf{a}$ and any final state \mathbf{b} , with $\mathbf{a}, \mathbf{b} \in \mathbf{R}^{n-|S|}$, there exists an input $\mathbf{x}_L(t)$ from the leader nodes S such that $\mathbf{x}_F(T) = \mathbf{b}$ for some finite T .

A system is *structurally controllable* if it is controllable for almost any choice of the weight matrix W . The following

theorem, proved in [9], relates the structural controllability of a leader-follower system to its graph topology.

Theorem 2: The leader-follower system with follower dynamics given by (1) is structurally controllable from the leader set S if and only if the following conditions hold. First, for any node $v \in V \setminus S$, there exists a directed path from some node $w \in S$ to v . Second, for any subset $A \subseteq V \setminus S$, $|A| \leq |N(A)|$. If the set S satisfies the first condition, the system is defined to be *accessible*. If the second condition is satisfied, then the system is *dilation-free*.

The relationship between the dilation-free property of Theorem 2 and graph matching theory is shown in the following lemma, which generalizes the minimum inputs theorem of [8].

Lemma 1: For a given leader set S , define a bipartite graph $H = (U_1, U_2, L)$ by $U_1 = N(V \setminus S)$, $U_2 = V \setminus S$, and $(i, j) \in L$ iff $(i, j) \in E$. Then the dilation-free property is satisfied by S iff there exists a perfect matching in H .

Proof: First, suppose that the system is dilation-free. Then for every $A \subseteq V \setminus S$, $|N(A)| \geq |A|$, and hence a perfect matching of the graph H exists by the Hall Marriage Theorem. Similarly, if such a matching m exists, then $|A| = |m(A)| \leq |N(A)|$. ■

Moreover, since we have assumed that the graph G is strongly connected, the accessibility criterion is satisfied for any S , and hence controllability from S is equivalent to the existence of a perfect matching from $N(V \setminus S)$ into $V \setminus S$ as in Lemma 1.

D. Matroid Theory and Submodularity

In what follows, if $v \in V$ and $A \subseteq V$, then $A + v$ is used to represent $A \cup \{v\}$. A matroid is then defined as follows.

Definition 2: A matroid \mathcal{M} consists of a finite set V and a collection \mathcal{I} of subsets of V satisfying the following properties: (i) $\emptyset \in \mathcal{I}$, (ii) if $B \in \mathcal{I}$, then $A \in \mathcal{I}$ for all $A \subseteq B$, and (iii) if $A, B \in \mathcal{I}$ and $|A| < |B|$, then there exists $x \in B$ such that $A + x \in \mathcal{I}$.

A maximal set in \mathcal{I} is denoted a *basis*. All bases of a matroid have the same cardinality; moreover, by property (ii) of Definition 2, any matroid can be characterized by its bases. A simple example of a matroid is the *uniform matroid*, $\mathcal{U}_{k,n}$, defined as the set of all subsets $S \subseteq V$, with $|V| = n$, satisfying $|S| \leq k$. The bases of this matroid are the sets with cardinality exactly k . The following lemma gives a composition rule for matroids [12].

Lemma 2: Let $\mathcal{M}_1 = (V, \mathcal{I}_1)$ and $\mathcal{M}_2 = (V, \mathcal{I}_2)$ be matroids. Then the set $\mathcal{M} = (V, \mathcal{I})$, with $A \in \mathcal{I}$ iff $A = A_1 \cup A_2$ for some $A_1 \in \mathcal{I}_1$ and $A_2 \in \mathcal{I}_2$, is a matroid.

Lastly, we define the properties of *submodularity* and *monotonicity* for set-valued functions [13].

Definition 3: Let V be a finite set. Then the function $f : 2^V \rightarrow \mathbf{R}$ is nondecreasing (resp. nonincreasing) if $A \subseteq B$ implies $f(A) \leq f(B)$ (resp. $f(A) \geq f(B)$). The function is strictly nondecreasing (nonincreasing) if $A \subsetneq B$ implies that the inequality is strict.

Definition 4: The function $f : 2^V \rightarrow \mathbf{R}$ is submodular if for any $A \subseteq B$ and $v \notin B$, $f(A + v) - f(A) \geq f(B + v) - f(B)$.

$f(B)$. f is supermodular if $-f$ is submodular.

Intuitively, submodularity is a notion of diminishing returns, analogous to concavity of continuous functions. It implies that adding a new element v to a relatively small set provides a larger incremental benefit than adding v to a larger set.

III. MATROID OPTIMIZATION FRAMEWORK

In this section, the problem of selecting leaders in order to optimize performance while guaranteeing controllability is presented within a matroid optimization framework. We first formulate a general matroid optimization problem for leader selection. We then give solution algorithms, followed by bounds on the accuracy of our algorithms that are satisfied when the objective function is supermodular.

A. Problem Formulation

We study the problem of selecting a set of leader agents S according to three criteria. First, the set of leader agents cannot exceed a fixed number k . Second, the follower agents should be controllable from the leader agents. Third, the leader agents should be chosen to minimize a cost function $f(S)$, provided that the first two constraints are satisfied. Letting

$$\mathcal{C} \triangleq \{S \subseteq V : V \text{ controllable from } S\}, \quad (2)$$

the problem can be formulated as

$$\begin{aligned} & \underset{S}{\text{minimize}} && f(S) \\ & \text{s.t.} && |S| \leq k \\ & && S \in \mathcal{C} \end{aligned} \quad (3)$$

The following theorem gives a formulation that is equivalent to (3).

Theorem 3: Define the set \mathcal{C}_k by

$$\mathcal{C}_k \triangleq \{S : S \in \mathcal{C} \text{ and } |S| = k\} \quad (4)$$

and define $\tilde{\mathcal{C}}_k \triangleq \{S : S \subseteq T \in \mathcal{C}_k\}$. If the function $f(S)$ is strictly decreasing as a function of S , then the set S^* solving the optimization problem

$$\begin{aligned} & \underset{S}{\text{minimize}} && f(S) \\ & \text{s.t.} && S \in \tilde{\mathcal{C}}_k \end{aligned} \quad (5)$$

is also a solution to (3).

Proof: Since $\mathcal{C}_k \subseteq \tilde{\mathcal{C}}_k$, it suffices to show that the optimal solution to (5) is a feasible set of (3). Since $f(S)$ is a strictly decreasing function of S , any solution to (3) and (5) must satisfy $|S| = k$. However, if $S \in \tilde{\mathcal{C}}_k$ and $|S| = k$, then $S \in \mathcal{C}_k$, since it would not be possible for S to be a proper subset of a set in \mathcal{C}_k . Hence the optimal solution to (5) is a feasible solution to (3). ■

Theorem 3 establishes that solving (3) is equivalent to solving the relaxed problem (5). The structure of (5) is given by the following theorem.

Theorem 4: The set $\mathcal{M}_k = (V, \tilde{\mathcal{C}}_k)$ is a matroid.

The following lemma gives a needed intermediate step to the proof.

Lemma 3: Define $k_0 \triangleq \min \{|S| : S \in \mathcal{C}\}$. Then \mathcal{M}_{k_0} is a matroid.

Proof: Define a bipartite graph $H = (N(V), V, E)$ as in Lemma 1. We first show that \mathcal{C}_{k_0} is equal to the set of subsets $S \subseteq V$ such that there exists a maximal matching $m : V \rightarrow V$ on H with all $v \in S$ unmatched. First, suppose that $S \in \mathcal{C}_{k_0}$. By Lemma 1, there exists a perfect matching $m : N(V \setminus S) \rightarrow V \setminus S$. Suppose that, for some $v \in S$, m can be extended to a matching $\tilde{m} : V \rightarrow V$ with $\tilde{m}(u) = v$ for some $u \in N(v)$. This implies that there is a perfect matching from $N(V \setminus (S - v))$ into $V \setminus (S - v)$, and hence the set $S - v$ is sufficient to control the network. However, $|S - v| = k_0 - 1$, contradicting the definition of k_0 . Therefore m is a maximal matching on V .

Now, suppose that there exists a maximal matching $m : V \rightarrow V$ such that all $v \in S$ are unmatched. Then m defines a perfect matching on $N(V \setminus S) \rightarrow V \setminus S$. Furthermore, since all maximal matchings have the same cardinality [10], $|V \setminus S| = n - k_0$, and hence $|S| = k_0$.

The above establishes that \mathcal{C}_{k_0} is equal to the set of subsets that are unmatched by any maximal matching. This family of sets, however, defines the dual of the *partition matroid*, and hence is a matroid [10, Lemma 3.1.5]. ■

Proof: [Proof of Theorem 4] Let k_0 be defined as in Lemma 3. Define \mathcal{C}'_k to be

$$\mathcal{C}'_k \triangleq \{A_1 \cup A_2 : A_1 \in \mathcal{C}_{k_0}, |A_2| \leq k - k_0\} \quad (6)$$

By Lemma 3, \mathcal{C}_{k_0} defines a matroid. It can also be shown that $\{S : |S| \leq k - k_0\}$ defines a matroid. Hence $\mathcal{M}' = (V, \mathcal{C}'_k)$ is a matroid by Lemma 2.

We now show that $\mathcal{C}'_k = \tilde{\mathcal{C}}_k$. Let \mathcal{B} and \mathcal{B}' be the sets of maximal sets in \mathcal{C}'_k and $\tilde{\mathcal{C}}_k$. Since \mathcal{C}'_k and $\tilde{\mathcal{C}}_k$ are equal to the sets of subsets of \mathcal{B} and \mathcal{B}' , respectively, it suffices to show that $\mathcal{B} = \mathcal{B}'$.

Suppose that S is a basis element of \mathcal{C}'_k . Then $S = B_1 \cup B_2$, where B_1 and B_2 are basis elements of \mathcal{C}_{k_0} and \mathcal{U}_{k-k_0} , respectively. Since $B_1 \subseteq S$ and $B_1 \in \mathcal{C}$, the system is controllable from S and hence $S \in \tilde{\mathcal{C}}_k$.

Now, let $S \in \mathcal{B}'$. Since $S \in \mathcal{B}'$, the system is controllable from S , and so by Lemma 1 there exists a perfect matching $m : N(V \setminus S) \rightarrow V \setminus S$. Now, m can be extended to a maximal matching $\tilde{m} : V \rightarrow V$ [10]. Let S_1 equal the set of nodes that are left unmatched by \tilde{m} , and let S_2 equal the set of nodes that are matched by \tilde{m} but not m . By Lemma 3, S_1 is a basis set of \mathcal{C}_{k_0} . Furthermore, since \tilde{m} has cardinality $n - k_0$ and m has cardinality $n - k$, $|S_2| = k - k_0$, and hence S_2 is a basis set of \mathcal{U}_{k-k_0} that is disjoint from S_1 . $S_1 \cup S_2$ is therefore a basis set of \mathcal{C}'_k . This establishes that the maximal sets of \mathcal{C}'_k and $\tilde{\mathcal{C}}_k$ are equal, completing the proof. ■

Together, Theorems 3 and 4 imply that selecting a set of up to k leaders in order to guarantee controllability while optimizing performance is a matroid optimization problem. Matroid optimization can be performed efficiently using greedy algorithms, as described in the following section.

B. Leader Selection Algorithms

A general algorithm for solving leader selection problems of the form (5) is as follows. Let S_t denote the leader set before the t -th iteration, with $S_0 = \emptyset$. At the t -th iteration, the first step is to find the set of nodes $V_t \triangleq \{v \in V \setminus S_{t-1} : S_{t-1} + v \in \tilde{C}_k\}$.

A procedure for determining whether $S_{t-1} + v \in \tilde{C}_k$ is as follows. The procedure assumes that a maximal matching m_{t-1} has been found such that S_{t-1} is unmatched. There are two cases. If v is unmatched by m_{t-1} , then $v \in V_t$. If v is matched, then a search is performed for a path $\pi = \{(v_0, w_0), (w_0, v_1), (v_1, w_1), \dots, (w_{r-1}, v_r)\}$ in the bipartite graph H , where $m_{t-1}(w_i) = v_i$, $(w_i, v_{i+1}) \in E$, $v_0 = v$, $v_i \notin S$, and v_r is unmatched by m_{t-1} . If such a path is found, then a new matching m_t^v is initialized with $m(w_i) = v_{i+1}$ for $i = 1, \dots, r-1$ and v is added to V_t . Otherwise v is not added to V_t . A proof of the correctness of this stage is given as Lemma 4 in the appendix. A recursive algorithm denoted **Determine- v - V_t** for finding such a path based on breadth-first search is given below.

Algorithm Determine- v - V_t : Determines if a node v is in the set V_t

Input: Graph topology $G = (V, E)$, set T , node v
Matching m that avoids T

Output: Matching m' that avoids $T + v$ if it exists
Otherwise outputs \emptyset

if v unmatched by m

return m

else

$w \leftarrow$ unique node satisfying $m(w) = v$

for all $j \in N(w) \setminus T$

$m_j \leftarrow$ Matching m restricted to $V \setminus \{w\}$

$\tilde{m}_j \leftarrow$ **Determine- v - V_t** ($G - w, T, j, m_j$)

if $\tilde{m}_j \neq \emptyset$

$m' \leftarrow \tilde{m}_j \cup \{(w, j)\}$

return m'

end if

end for

end if; return \emptyset

Among the nodes in V_t , let v_t^* denote the node such that $f(S_{t-1} + v_t^*)$ is minimized. Set $S_t = S_{t-1} + v_t^*$ and increment t . The algorithm terminates when $t = k$. A pseudocode description of the algorithm is denoted **k -leaders-matroid**.

Bounds on the optimality of the leader set returned by algorithm **k -leaders-matroid** can be derived for certain properties of the function $f(S)$, as described by the following theorem.

Theorem 5: Let $\hat{f} = \max_{v \in V} f(\{v\})$, and let f^* be the optimum value of $f(S)$ in (3). If $f(S)$ is strictly decreasing and supermodular as a function of S , then the set S^* returned by **k -leaders-matroid** satisfies

$$f(S^*) \leq \frac{1}{2}f^* + \frac{1}{2}\hat{f} \quad (7)$$

Proof: Since $f(S)$ is strictly decreasing as a function of S , (3) and (5) are equivalent (Theorem 3). In [14] it was

Algorithm k -leaders-matroid: Algorithm for selecting a set of k leaders that satisfy controllability while minimizing objective f

Input: Graph topology $G = (V, E)$,

maximum number of leaders k

Output: Set of leaders S

Initialization: $S_0 \leftarrow \emptyset$, $t \leftarrow 1$

$m_0 \leftarrow$ maximal matching of V into V

while $t \leq k$

$V_t \leftarrow \emptyset$

for $v \in V \setminus S$

$m_t^v \leftarrow$ **Determine- v - V_t** (G, S_{t-1}, m_{t-1})

if $m_t^v \neq \emptyset$: $V_t \leftarrow V_t + v$

end for $v^* \leftarrow \arg \min \{f(S_{t-1} + v) : v \in V_t\}$

$S_t \leftarrow S_{t-1} + v^*$

$m_t \leftarrow m_t^v$

end while

$S \leftarrow S_t$, **return** S

shown that, for a matroid \mathcal{I} and monotone nonincreasing supermodular function $h(S)$, the set S formed by selecting at each iteration the node v defined by

$$v = \arg \min_v \{h(S + v) : (S + v) \in \mathcal{I}, v \notin S\},$$

satisfies

$$h(S) \leq \frac{1}{2}h(S^*) + \frac{1}{2}\hat{h} \quad (8)$$

where S^* is the optimal set and \hat{h} is the maximum value of h . By Theorem 4, \tilde{C}_k is a matroid, and hence the algorithm **k -leaders-matroid** satisfies (8). ■

Some examples of supermodular objective functions $f(S)$ are as follows. In [6], it was shown that, when the followers estimate the leader state using a best linear estimator, the resulting mean-square error in steady-state is a supermodular function of S . Similarly, the network coherence, defined as the asymptotic deviation of the node states from their consensus value, was shown in [2] to be proportional to $\text{tr}(L(S)^{-1})$, where $L(S)$ is obtained by removing the rows and columns with indices in S from the graph Laplacian L . By the argument of [6], this function is supermodular as well. Furthermore, in [5], it was proved that the l^p -norm of the deviation of a node's state from its desired final state is a supermodular function of S . Leaders can be selected to optimize any of these performance metrics while guaranteeing controllability and a provable optimality bound as a result of Theorem 5.

If an additional linearity constraint is imposed on $f(S)$, then **k -leaders-matroid** can be shown to be optimal.

Theorem 6: If the function $f(S) = \sum_{i \in S} w_i$ for some $w_i \leq 0$, then **k -leaders-matroid** returns the solution set S^* to problem (3).

Proof: Theorem 1 of [12, Chapter 19] states that an optimal set S that minimizes an objective function $\sum_{i \in S} w_i$, where the w_i 's are constant and S belongs to a matroid, can be found by a greedy algorithm. Theorem 6 then follows from the fact that (3) and (5) are equivalent, as well as the

fact that $\tilde{\mathcal{C}}_k$ is a matroid. The condition $w_i \leq 0$ is needed to ensure that the algorithm does not terminate until $|S| = k$; otherwise, the returned set S might not be a basis set of $\tilde{\mathcal{C}}_k$, and hence controllability cannot be guaranteed. ■

IV. LEADER SELECTION VIA SUBMODULAR RELAXATION

The problem of selecting a set S of up to k leaders in order to ensure controllability and minimize a cost metric $f(S)$ implicitly assumes that a fixed number of nodes k is sufficient to control the network. When this is not the case, a relaxed formulation is needed to ensure that as much of the network as possible is controllable. The relaxed controllability criterion should be a real number in the interval $[0, 1]$, taking value 1 if the system is controllable and 0 if no nodes can be controlled. We first define the function $c(S)$ for a leader set S as

$$c(S) \triangleq \max \{|V'| : (V', E') \subseteq G \text{ is controllable from } S\}.$$

The function $c(S)$ is equal to the size of the largest subgraph of G that is controllable from the leader set S . We define the term *graph controllability index* (GCI) to be equal to $\frac{c(S)}{n}$, the fraction of nodes that can be controlled from S , which is equal to 1 when the entire network is controllable. Based on this definition, a relaxed version of the problem of selecting up to k leaders to guarantee controllability is given by

$$\begin{aligned} & \text{maximize} && \frac{1}{n}c(S) - \lambda f(S) \\ & S && \\ & \text{s.t.} && |S| \leq k \end{aligned} \quad (9)$$

where λ is a nonnegative constant. Under the formulation (9), a set of leaders is chosen in order to maximize controllability, subject to a penalty introduced by the cost function $f(S)$. If $c(S)$ is submodular as a function of S and $f(S)$ is supermodular as a function of S , then $\frac{1}{n}c(S) - \lambda f(S)$ is submodular, leading to efficient algorithms for solving (9). The following theorem establishes the submodularity of $c(S)$.

Theorem 7: $c(S)$ is a submodular function of S .

Proof: By Lemma 1, a subgraph $G' = (V', E')$ is controllable from S if there exists a perfect matching from $N(V' \setminus S)$ into $V' \setminus S$. A maximum controllable subgraph is therefore given by the set of nodes matched by a maximal matching of $N(V \setminus S)$ into $V \setminus S$, denoted $g(S)$. The function $c(S)$ is then equal to $g(S)$, plus the size of the set S itself, so that $c(S) = |S| + g(S)$.

Since $|S|$ is a submodular function of S [15], it suffices to show that $g(S)$ is submodular. Let $S \subseteq T$, and suppose $v \notin T$. The function $g(S)$ is submodular if and only if

$$g(S+v) - g(S) \geq g(T+v) - g(T). \quad (10)$$

We now analyze the possible values of $g(T+v) - g(T)$. If there is a maximal matching of $N(V \setminus S)$ into $V \setminus S$ that avoids v , then $g(S+v) = g(S)$. Otherwise, if every maximal matching of $N(V \setminus S)$ into $V \setminus S$ includes v , then removing v from $V \setminus S$ reduces the cardinality of every matching by 1. Hence $g(S+v) = g(S) - 1$ in this case.

In order to prove (10), we observe that, by the preceding analysis, $(g(T+v) - g(T)) \in \{-1, 0\}$ and $(g(S+v) - g(S)) \in \{-1, 0\}$. If $g(T+v) - g(T) = -1$, then (10) is satisfied automatically, since $g(S+v) - g(S) \geq -1$. If $g(T+v) - g(T) = 0$ holds, then it must be shown that $g(S+v) = g(S)$. This is equivalent to proving that if there is a maximal matching of $N(V \setminus T)$ into $V \setminus T$ that avoids v , then there is a maximal matching of $N(V \setminus S)$ into $V \setminus S$ that avoids v as well.

Let m be a maximal matching of $V \setminus T$ such that v is unmatched, and let m' be a maximal matching of $N(V \setminus S)$ into $V \setminus S$. If v is left unmatched by m' , then $g(S) = g(S+v)$ and (10) holds. Otherwise, we give a sequential procedure for constructing a maximal matching \tilde{m} of $N(V \setminus S)$ into $V \setminus S$ such that v is unmatched.

The first step of the procedure is as follows. Define $v_0 = v$ and let u_0 satisfy $m'(u_0) = v_0 = v$. Since u_0 is a neighbor of $v_0 \in V \setminus T$, $u_0 \in N(V \setminus T)$. If u_0 is unmatched by m , then m can be extended by setting $m(u_0) = v$, contradicting the assumption that m is maximal. Define $v_1 = m(u_0)$. We have that $v_1 \neq v$, since v_1 is matched under m and v is not.

At step l in the procedure, where $l > 1$, l distinct elements of $N(V \setminus T)$, denoted u_0, \dots, u_{l-1} , are given, along with $(l+1)$ distinct elements of $V \setminus T$, denoted v_0, \dots, v_l . The elements satisfy $m'(u_r) = v_r$ and $m(u_r) = v_{r+1}$ for $r = 0, \dots, l-1$. There are two cases. In the first case, v_l is unmatched by m' . In the second case, v_l is matched by m' .

Case I – v_l is unmatched by m' : If v_l is unmatched by m' , then \tilde{m} can be constructed by setting $\tilde{m}(u_r) = v_{r+1}$ for $r = 0, \dots, l-1$ and $\tilde{m}(w) = m'(w)$ for $w \notin \{u_0, \dots, u_r\}$. The matching \tilde{m} has the same cardinality as m' , and hence is maximal. Moreover, it avoids v , and hence the procedure returns \tilde{m} and terminates.

Case II – v_l is matched by m' : In this case, there exists u_l with $m'(u_l) = v_l$. Since $v_l \in V \setminus T$, the element u_l is in $N(V \setminus T)$. Assume that v_l is left unmatched by m . The cardinality of m could then be strictly increased by setting $m(u_r) = v_r$ for $r = 0, \dots, l$ (since m would match all the same elements of $V \setminus T$ in addition to v). This, however, contradicts the assumption that m is maximal. Hence $m(u_l) = v_{l+1}$ for some $v_{l+1} \in V \setminus T$. Furthermore, the elements $u_0, \dots, u_l \in N(V \setminus T)$ are distinct, since they are matched by m' to v_0, \dots, v_l , which are distinct by inductive hypothesis. The elements $v_0, \dots, v_{l+1} \in V \setminus T$ are distinct since they are matched by m to distinct elements u_0, \dots, u_l . The value of l is incremented to $(l+1)$ and the procedure continues.

By construction, this procedure terminates iff a maximal matching $\tilde{m} : N(V \setminus S) \rightarrow V \setminus S$ that avoids v is found. To complete the proof, it suffices to show that the procedure does in fact terminate. Suppose that $N(V \setminus T)$ has cardinality K , and suppose that the procedure has reached step K . Then $(K+1)$ distinct elements $u_0, \dots, u_K \in N(V \setminus T)$ have been found, contradicting the assumption that there are only K distinct elements in $N(V \setminus T)$. The procedure therefore terminates within K steps by returning a matching $\tilde{m} : N(V \setminus S) \rightarrow V \setminus S$ that avoids v , as desired. ■

Theorem 7 implies that (9) is a submodular optimization problem when the function $f(S)$ is supermodular as a function of S . Since $\frac{1}{n}c(S) - \lambda f(S)$ is submodular, efficient algorithms can be designed that jointly maximize controllability and performance, as described in the following section.

A. Leader Selection Algorithms for Relaxed Formulation

An algorithm for selecting a set of leaders S under the relaxed formulation of (9), in which leaders are selected to maximize the largest controllable subgraph of G , is as follows. The leader set is initialized to $S = \emptyset$. At the t -th iteration, the node $v_t \in V \setminus S$ that maximizes $\frac{1}{n}c(S) - \lambda f(S)$ is selected and S is updated to $S \cup \{v_t\}$. In other words, at each iteration, the next best possible leader is added to the leader set. The algorithm terminates after k iterations, where k is the maximum number of leaders. A pseudocode description is given as **k -leaders-relaxed**.

Algorithm k -leaders-relaxed: Algorithm for selecting up to k leaders to maximize controllability while minimizing the cost $f(S)$

Input: Graph $G = (V, E)$,

maximum number of leaders k

Output: Leader set S

Initialization: $t \leftarrow 0$, $S \leftarrow \emptyset$

while $t < k$

$v \leftarrow \arg \max_{v \in V \setminus S} \left\{ \frac{1}{n}(c(S+v) - c(S)) - \lambda(f(S+v) - f(S)) \right\}$

$S \leftarrow S + v$, $t \leftarrow t + 1$

end while

return S

Theorem 8: Suppose that the function f is a monotone supermodular function of S . Let \hat{S} be the solution to (9), and let S^* be the set returned by **k -leaders-relaxed**. Then

$$\frac{1}{n}c(S^*) - \lambda f(S^*) \geq \left(1 - \frac{1}{e}\right) \left[\frac{1}{n}c(\hat{S}) - \lambda f(\hat{S}) \right] \quad (11)$$

Proof: Since $f(S)$ is supermodular, $-\lambda f(S)$ is submodular, and so $\frac{1}{n}c(S) - \lambda f(S)$ is submodular. Hence the theorem holds by Theorem 4.1 of [13], which establishes that, for a monotone submodular function, the greedy algorithm returns a set that is within a factor of $(1 - 1/e)$ of the optimum value. ■

Theorem 8 establishes that the algorithm **k -leaders-relaxed** that are within a provable bound of the optimum. Moreover, if $\lambda = 0$, corresponding to the case where leaders are selected based on controllability alone, the algorithm returns a minimum-size set of leaders S such that G is controllable from S . In other words, when $\lambda = 0$, **k -leaders-relaxed** is optimal.

Theorem 9: If $\lambda = 0$ and there exists a set S with $|S| \leq k$ such that G is controllable from S , then G is controllable from the set returned by **k -leaders-relaxed**.

Proof: Let k_0 be the minimum number of agents needed to control the system, as in 3, so that $c(\emptyset) = n - k_0$. At each iteration, a node that maximizes $c(S+v) - c(S) =$

$|S+v| - |S| + (g(S+v) - g(S))$ is added to S . Suppose that, at each iteration, a node v is added such that $g(S+v) = g(S)$. Then since $c(S)$ is incremented by 1 at each iteration and $c(S) = n - k_0$ initially, $c(S) = n$ after k_0 iterations, implying that a controllable leader set is chosen if it exists.

It remains to show that, at each iteration, there exists $v \in V \setminus S$ such that $g(S+v) = g(S)$. Suppose this is not true when $|S| < k_0$. Then for any $v \in V \setminus S$ and for any maximal matching m , $m(u) = v$ for some $u \in N(V \setminus S)$. This, however, implies that there is a maximal matching with unmatched set equal to S , contradicting the assumption that $|S| < k_0$. Hence S is insufficient to control the system. ■

B. Example of Joint Leader Selection

As an example, consider the system with network topology given by the graph G in Figure 1(a). We analyze leader selection for this network under the two criteria of controllability and coherence. Coherence is optimized by minimizing the deviation $f(S) = \text{tr}((L(S))^{-1})$, where L is the Laplacian matrix of the graph and $L(S)$ is the matrix obtained by removing the rows and columns with indices in S from L [7]. Note that, while an undirected graph is used to ensure that the coherence metric is well-defined, our approach is valid for both directed and undirected graphs for other metrics.

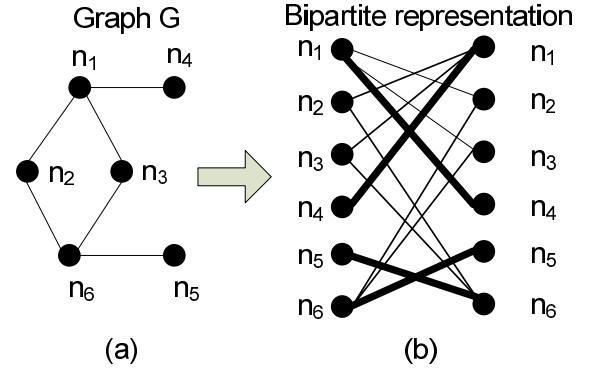


Fig. 1. Illustration of the relationship between matching and our proposed framework. (a) An example network topology. (b) The equivalent bipartite representation. Each line represents an edge in the graph G . The bold lines represent one possible maximal matching. The leader set corresponding to this maximal matching is equal to the set of unmatched nodes $\{n_2, n_3\}$.

The matching-based selection algorithm of [8] chooses a leader set by identifying a maximal matching of the associated bipartite graph (Figure 1(b)). A matching can be found, for example, by using the Edmonds matching algorithm [10, Ch 1]; one possible matching is given by the bold lines in Figure 1(b). Since neither n_2 nor n_5 can be added to the matching without removing an existing node, this matching is maximal. Moreover, these unmatched nodes correspond to the leader set, with associated network coherence of 4.

An analysis based on the joint selection approach of algorithm **k -leaders-relaxed** is as follows, with $\lambda = 0.1$, and $k = 2$. Initially, there are four nodes that can be left unmatched by a maximal matching on G , namely $\{n_2, n_3, n_4, n_5\}$. These nodes can be identified, for example, by using algorithm

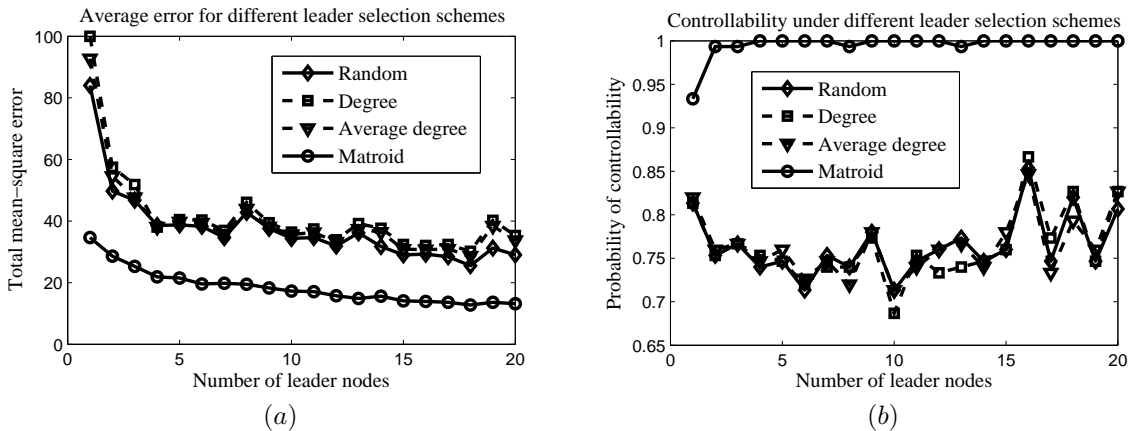


Fig. 2. Comparison of our proposed matroid optimization approach with other leader selection algorithms. (a) Our proposed method achieves significantly lower mean-square error in steady-state than other algorithms. Note that random leader selection outperforms degree-based selection. (b) The matroid optimization approach achieves controllability for all chosen networks, while the other methods do not guarantee controllability in general.

Determine- v - V_t in Section III-B. Of these nodes, n_2 and n_3 result in the minimum deviation of 5.0. For this example, suppose that $S = \{n_2\}$ is chosen at the first iteration. Choosing n_1 or n_6 will not cause an incremental increase in $c(S)$, and hence are suboptimal.

At the second iteration, a node v is chosen to maximize $c(\{2\}+v) - \lambda f(\{2\}+v)$. Only n_3 , n_4 , and n_5 are unmatched by a maximal matching on $V \setminus S$. Of these two nodes, n_3 leads to a deviation $f(S)$ of 3.0, and hence is chosen, for a leader set of $S = \{n_2, n_3\}$. As in the algorithm of [8], this leader set also leads to controllability. In addition, however, there is a 25% reduction in incoherence.

V. SIMULATION

We evaluated the performance of the matroid optimization approach of Section III via Matlab simulation study. We generated an Erdos-Renyi random graph $G(n, p)$ with $n = 100$ total nodes [16]. The probability that edge (i, j) exists was set to 0.05, with the probability of (i, j) existing independent of (i', j') existing for all $i \neq i'$ and $j \neq j'$. Each link (i, j) was assumed to experience additive white Gaussian noise with variance ν_{ij}^2 . The variance ν_{ij}^2 was chosen at random from an exponential distribution with mean 1. Each data point represents an average over 150 independent trials.

Leader selection was carried out using four algorithms. Under the first algorithm, a set of k leaders was selected uniformly at random. In the second algorithm, the k nodes with highest degree were selected as leaders. Under the third algorithm, the k nodes with average degree were chosen as leaders. The fourth algorithm was the matroid optimization approach defined in the k -**leaders-matroid** algorithm. The mean-square error of the nodes in steady-state, as defined in [6], was chosen as a performance metric.

Figure 2(a) shows the level of error achieved for each leader selection approach. The matroid optimization approach provides significantly lower average error in steady-state than the other methods. As an example, selecting 5 leader nodes using the matroid-based approach leads to a mean-square deviation from the consensus value of 20 in

steady-state, compared with 40 for the degree-based and random methods. This suggests that incorporating controllability as a design constraint does not lead to substantially lower performance. Furthermore, the random selection method slightly outperforms selection based on average degree, which in turn gives lower error than choosing high-degree nodes.

The controllability of the system under each algorithm is illustrated in Figure 2(b). In the figure, each data point represents the fraction of the trials that resulted in a controllable network using the specified algorithm. For the matroid-based approach, the probability of controllability was initially less than 1 for small values of k , since for some networks it is not possible to achieve controllability with only one or two leaders. As k increases, the probability of controllability for the matroid-based approach converges to 1.

The random and degree-based algorithms result in a probability of controllability that varies between 0 and 1 but does not converge to 1. In the case of degree-based selection, this is supported by earlier work of [8], which suggests that highly-connected, high-degree nodes may be unsuitable for controlling a network. The fact that typically leaders chosen using random selection are unable to control the network follows from the fact that only a relatively small subset of nodes are capable of satisfying the controllability criterion when chosen as leaders, and that these nodes are not likely to be simultaneously selected by random chance. This further reinforces the need for unified framework for leader selection that incorporates controllability as a design constraint while meeting performance requirements.

VI. CONCLUSIONS AND FUTURE WORK

Leader selection in MAS can be performed according to two design approaches. Under the first approach, a fixed number of leaders are chosen to achieve controllability while optimizing a performance metric. In the second approach, the minimum-size set of leaders is chosen in order to satisfy controllability while meeting a required bound on a performance metric.

In this paper, we studied the problem of selecting a fixed number of leaders based on joint consideration of controllability and performance metrics. We showed that the joint formulation can be studied within a matroid optimization framework, in which controllability represents a matroid constraint. This formulation leads to algorithms with provable accuracy bounds for the case where the performance metric is a supermodular function of the leader set, and optimal algorithms when the metric is a linear function of the leader nodes.

For the case where the number of leaders is insufficient to guarantee controllability, or where performance criteria such as robustness to noise are given equal weight to controllability, we introduce a relaxed formulation based on the fraction of nodes that are controllable from the leader set. We then proved that this fraction of nodes is a submodular function of the leader set, leading to efficient leader selection algorithms. Our theoretical results for the matroid optimization approach presented in this paper were also illustrated and supported through example and simulations.

In our future work, we will study the case where the number of nodes that can act as leaders is not fixed. In this case, a bound on the system performance is given and the leader set must be chosen in order to achieve controllability and satisfy this bound. We plan to investigate this leader selection problem within both the matroid optimization and submodular relaxation frameworks.

REFERENCES

- [1] P. Wang and F. Hadaegh, "Coordination and control of multiple microspacecraft moving in formation," *Journal of the Astronautical Sciences*, vol. 44, no. 3, pp. 315–355, 1996.
- [2] S. Patterson and B. Bamieh, "Leader selection for optimal network coherence," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 2692–2697.
- [3] H. Tanner, "On the controllability of nearest neighbor interconnections," in *Proceedings of the 43rd IEEE Conference on Decision and Control*, vol. 3. IEEE, 2005, pp. 2467–2472.
- [4] S. Jafari, A. Ajorlou, A. Aghdam, and S. Tafazoli, "On the structural controllability of multi-agent systems subject to failure: a graph-theoretic approach," in *Proceedings of the 49th IEEE Conference on Decision and Control (CDC)*. IEEE, 2010, pp. 4565–4570.
- [5] A. Clark, L. Bushnell, and R. Poovendran, "Leader selection for minimizing convergence error in leader-follower systems: A supermodular optimization approach," in *Proceedings of the 10th IEEE International Symposium on Modeling and Optimization in Mobile, Ad Hoc, and Wireless Networks (WiOpt)*, 2012.
- [6] A. Clark and R. Poovendran, "A submodular optimization framework for leader selection in linear multi-agent systems," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. IEEE, 2011.
- [7] F. Lin, M. Fardad, and M. Jovanovic, "Algorithms for leader selection in large dynamical networks: Noise-corrupted leaders," in *Proceedings of the 50th IEEE Conference on Decision and Control and European Control Conference (CDC-ECC)*. IEEE, 2011.
- [8] Y. Liu, J. Slotine, and A. Barabási, "Controllability of complex networks," *Nature*, vol. 473, no. 7346, pp. 167–173, 2011.
- [9] C. Lin, "Structural controllability," *Automatic Control, IEEE Transactions on*, vol. 19, no. 3, pp. 201–208, 1974.
- [10] L. Lovász and M. Plummer, *Matching theory*. Elsevier Science Ltd, 1986, no. 121.
- [11] C. Chen, *Linear System Theory and Design*. Oxford University Press, 1976.
- [12] D. Welsh, *Matroid theory*. Academic Press London, New York, San Francisco, 1976, vol. 5.

- [13] G. Nemhauser, L. Wolsey, and M. Fisher, "An analysis of approximations for maximizing submodular set functions - I," *Mathematical Programming*, vol. 14, no. 1, pp. 265–294, 1978.
- [14] M. Fischer, G. Nemhauser, and L. Wolsey, "An analysis of approximations for maximizing submodular set functions-II," *Mathematical Programming Studies*, vol. 8, pp. 73–87, 1978.
- [15] S. Fujishige, *Submodular functions and optimization*. Elsevier Science, 2005, vol. 58.
- [16] B. Bollobás, *Random graphs*. Cambridge University Press, 2001, vol. 73.

APPENDIX

The following lemma proves that the algorithm **Determine- $v-V_t$** of Section III-B correctly determines whether $v \in V_t$.

Lemma 4: Let $G = (V, E)$ be a graph, and let m be a maximal matching on G such that m avoids a set T . Then there exists a maximal matching m' on G that avoids $T + v$ if and only if there exists a path $\pi = \{(v_0, w_0), (w_0, v_1), (v_1, w_1), \dots, (w_{r-1}, v_r)\}$ in the bipartite graph H , where $m(w_i) = v_i$, $(w_i, v_{i+1}) \in E$, $v_0 = v$, $v_i \notin S$, and v_r is unmatched by m .

Proof: Suppose that the matching m avoids v . Then let $m' = m$, and there is a trivial path $\pi = \{v\}$. Now, suppose that $m(w) = v$ for some w . First let m' be a maximal matching that avoids $T + v$. Choose $w_0 = w$, and let $v_1 = m'(w_0)$. If $v_0 = v_1 = v$, then a path π is defined by $\pi = \{(v_0, w_0)\}$, as desired.

Otherwise, at each iteration, after v_i has been selected, choose w_i such that $m(w_i) = v_i$. If no such w_i exists, then $\pi = \{(v_0, w_0), (w_0, x_1), \dots, (w_{i-1}, v_i)\}$ is a path satisfying the conditions of the lemma. On the other hand, if such a w_i exists, then let $v_{i+1} = m'(w_i)$ (if w_i were unmatched, then a larger matching could be derived by setting $m'(w_i) = v_i$ and leaving the rest of m' unchanged, contradicting the assumption that m' is maximal).

Hence, at each iteration, either a path π is found, or a pair of edges (v_i, w_i) and (w_i, v_{i+1}) are found with $v_i \neq v_j$ for $i \neq j$, $w_i \neq w_j$ for $i \neq j$, and $v_i \notin T + v$. Since the v_i 's are distinct, the process must terminate within $n - (|T + v|)$ iterations; when termination occurs, a path π satisfying the conditions of the lemma has been found.

Conversely, suppose that such a path π has been found. Then a new matching m' that avoids $T + v$ can be found by setting $m'(w_i) = v_{i+1}$ for all w_i and $m'(w) = m(w)$ for all other $w \in N(V \setminus T)$. ■