Optimal Spread in Network Consensus Models

Fern Y. Hunt
Applied and Computational Mathematics Division
National Institute of Standards and Technology
Gaithersburg, Maryland 20899

February 3, 2014

Abstract

We consider the problem of identifying a subset of nodes in a network that will enable the fastest spread of information in a decentralized communication environment. In a model of communication based on a random walk on an undirected graph, the optimal set over all sets of the same or smaller cardinality, minimizes the sum of the mean first arrival times to the set by walkers starting at nodes outside the set. This problem was previously considered in [1] and in the context of consensus models in [3]. The set function $F$ to be minimized is supermodular and therefore the greedy algorithm is a principal tool used for constructing optimal sets or their approximations. However the resulting set is not always optimal and the degree of optimality although guaranteed is limited by the factor $(1 - 1/e)$.

In this paper, the problem is reformulated by limiting the scope of sets to be considered based on specific structural features of the graph. We seek solutions of the optimization problem or approximations of solutions with cardinality $p < K$ that are subsets of solutions of the optimization problem for cardinality $K$. Our main result is a theorem that gives sufficient conditions for a vertex cover to contain subsets that are optimal or near optimal approximations. Here optimality of a set is defined in terms of its $F$ value relative to the $F$ value of the solution of the problem for cardinality $K$. Each optimal or near optimal subset can be constructed element by element. Optimal and near optimal sets are solution candidates for the original problem.

Although we do not specify an algorithm employing this methodology, greedoid theory is used to construct a greedoid graph whose nodes are optimal and near optimal feasible sets. This structure is amenable to the stepwise construction of approximations and local search for improvements by a sequence of moves to “adjacent” nodes. Thus this paper suggests that the connection between the network graph topology and the topology of the greedoid graph is a useful line of inquiry in the search for new approaches to the problem of optimal spread.

1 Random Walk Consensus Model

Given a connected graph $G = (V, E)$, with vertices or nodes $V$ and edges $E$, we imagine a random walker situated at a node $i \in V$, moving to another node $j \in V$ in a single discrete time step. The choice of $j$ is random and has probability,

$$\text{Prob}\{i \to j\} = \begin{cases} p_{ij} > 0, & \text{if } (i,j) \in E \\ p_{ij} = 0 & \text{otherwise.} \end{cases}$$

The matrix $P = (p_{ij})_{i,j=1\ldots N}$ is the transition matrix of a Markov chain which in this paper, is assumed to be irreducible and aperiodic ([7]). $N$ is the number of nodes and as in [1] the spread
of information is described in terms of a process that is dual to the movement from informed to uninformed nodes. A random walk begins outside a pre-determined set $A$ of informed target nodes and ends at $A$. Starting at node $i \notin A$, a random walker first reaches the set $A$ at a hitting time $T_A = \min\{n > 0 : X_n \in A\}$, where $X_n$ is the node occupied by the walker at time $n$. The hitting time is closely related to the rate of convergence in a leader-follower model of Clark et al, as well as other consensus models \cite{3,9}. The effectiveness of a set $A$ in the spread of information by random walks can then be measured by,

$$F(A) = \left(\sum_{i \notin A} h(i, A)\right),$$

where $h(i, A) = E_i[T_A]$, is the expected number of steps to $A$ starting at node $i$. When $F(A)$ is small, $A$ is a desirable choice of informed nodes but is a poor choice if $F(A)$ is large. A standard result in Markov chain theory tells us that $h(i, A)$ is the $i$th component of the vector $H$, which solves the linear equation.

$$H = 1 + P_A H$$

where $1$ is a column vector of $N - |A|$ ones and $P_A$ is the matrix that results from crossing out the rows and columns of $P$ corresponding to the nodes of $A$ \cite{7}. Limited resources can constrain the maximum size of the subset to be selected so it makes sense then to ask for the most effective spreader subject to a cardinality constraint, i.e.

$$\min_{A \subset V, |A| \leq M} F(A)$$

Borkar et al \cite{11} showed that for arbitrary subsets $A, B \subseteq V$, $F(A \cap B) + F(A \cup B) \geq F(A) + F(B)$, that is, $F$ is a supermodular function. Clark et al in \cite{3} discussed a continuous time leader follower problem where a set $A$ of leader nodes are assigned fixed function values and the remaining follower nodes update their function values by weighted exchange with their neighbors as defined by equation (1). If $G$ is strongly connected, the node function values converge to a consensus value (vector) defined by the leader nodes and the rate of convergence has a connection to precisely the random walk problem we are describing. In \cite{3}, two leader selection optimization problems are posed. The first, is to select up to $M$ leaders in order to minimize the convergence error. Given the random walk connection the problem in Clark is equivalent to the the problem posed in equation (4).

Since $-F$ is submodular, both works make use of the work of \cite{10} to devise a greedy algorithm that builds an approximate solution to the optimization problem (4) in a stepwise fashion until a set of cardinality $M$ is reached. Recall that at the first stage of construction, the node with the smallest value of $F$ among the nodes is selected. At the $p$th stage the node is added to the set that results in a set of cardinality $p + 1$ with the smallest value of $F$. Using the results of \cite{10}, Borkar et al were able to give some guarantee of the quality of the approximation of a weaker version of the optimization problem where the additional constraint that the set contain an element $\{m\}$ is added. If $F^*_m$ is the optimal value of this weaker optimization problem and $A_K(m)$ is the greedy approximation obtained by starting the algorithm with the set $\{m\}$, then

$$F(A_K(m)) \leq (1 - \frac{1}{e})F^*_m + \frac{1}{e}F(\{m\})$$

In the continuous time setting, Clark et al obtained a similar inequality but it is independent of the choice of a required initial element.

It will be instructive in what follows to see the results of applying the greedy algorithm to some specific graphs. In all the examples discussed in this paper we assume that for every neighbor $j$ of
node \( i \), \( p(i,j) = \frac{1}{\deg(i)} \).

**EXAMPLE 1** Figures 1-4 illustrate a graph for which the greedy algorithm starting with the vertices 3 or 4 can generate optimal subsets up to \( K=4 \) as shown as well as \( K=5 \) (not shown).

**EXAMPLE 2** Figures 5-9 show a graph where the greedy algorithm does not generate optimal sets starting from a single vertex.

In this paper, we will reformulate the optimization problem with a view towards improving the limitations proved in [1], [3]. In section 2 we seek solutions of problem (4) that are subsets of maximal matches. These are optimal sets have cardinality larger than \( M \) (see section 2.1). Unfortunately it is not always possible to find optimal subsets so we enlarge our search to the class of optimal and near optimal sets of a specified degree of optimality in the sense of (9) (see section 2.2). Using a result in [5] we show that the enlarged class contains a subclass of sets that are the feasible sets of a greedoid. In particular this means that feasible optimal and near optimal sets can be constructed stepwise in an element by element fashion. Moreover feasible vertex covers that are large enough contain optimal or nearly optimal sets with a pre-specified degree of optimality. Our purpose here is not to present a specific algorithm. That is a topic of future research. This paper shows the existence of a structure of optimal and non-optimal sets which is amenable to algorithms based on approximation and improvements using local search. In section 3 results of the paper are summarized and future research is discussed.

## 2 Finding and Approximating Optimal Sets

### 2.1 Maximal Matches

The optimization problem as posed in equation (4) assumes no advance knowledge about the optimal set or any other possibly related sets. In this section we seek to explore alternative formulations of the problem that could lead to other possibly better approximations of the optimal set. We consider a process of obtaining optimal sets as subsets of existing ones. The next definition will be helpful in the discussion that follows:

**Definition 1** A vertex cover of a graph \( G = (V,E) \) is a set of vertices that are incident to every edge in \( E \).

Let \( A \) be a vertex cover. Since every edge is incident to an element of \( A \), a random walker starting at a vertex \( i \) outside of \( A \) must hit \( A \) at the first step. That is \( h(i,A) = 1 \). Now equation (3) implies that \( h(i,A) \geq 1 \) so it follows that \( A \) must be an optimal set for its own cardinality. Thus the problem of getting an optimal subset is partially resolved if one can construct a vertex cover. Fortunately, there is a simple greedy algorithm (sometimes called the Two Opt algorithm) for constructing a special case. The algorithm was discovered independently by Gavril and Yannakakis. The actual algorithm can be found in [4].

**Definition 2** A maximal match of a graph is a set of edges that are non-adjacent. The set is maximal in the sense that there is no larger set with this property.

As is well known, [6], the vertices of a maximal match form a vertex cover. To see why note that every edge \( e \) in \( E \) is either an edge of a maximal match or is adjacent to such an edge. Thus \( e \) contains a vertex in the match. That is, \( e \) is incident to some vertex in the match so the definition of
vertex cover is satisfied. Now let $\mathcal{M}$ be the set of vertices of the maximal match that was constructed using the Two Opt algorithm. It can be shown (3) that,

$$OPT \leq |\mathcal{M}| \leq 2 \cdot OPT,$$

where $OPT$ is the cardinality of the minimal vertex cover. The run time of Two Opt is $O(|E|)$ [4]. Because supersets of a vertex cover are optimal sets and since we observe that optimal sets are often subsets of a vertex cover, it is natural to seek solutions of the problem among the subsets of $\mathcal{M}$:

$$\min_{A \subseteq \mathcal{M}, |A| \leq M} F(A).$$

Recall that the graph in EXAMPLE 2 has optimal sets which cannot be found from application of the greedy algorithm. When the Two Opt algorithm is applied one obtains edges of a maximal match $(1, 3), (5, 6), (7, 8)$, whose vertices $M$, contain optimal subsets for $K = 1, 2, 3, 4, 6$. Figures 6-10 show the optimal sets in EXAMPLE 2 for $K = 1 - 6$. In contrast to the greedy approach of building up to an optimal set, we start with a maximal match and obtain optimal sets of smaller cardinality as subsets. Unfortunately this approach is not always successful.

EXAMPLE 3:
The vertices of two maximal matches for a graph are shown in Figure 11. The figure on the left shows a maximal match whose vertex set $\mathcal{M}$ contains no optimal subsets except itself while the match shown on the right contains subsets that are optimal sets $K = 1$ through 4. Both maximal matches were obtained using the Two Opt algorithm.

Do we have a way to predict when optimal sets of predetermined cardinality are contained in a maximal match? Presently we do not. The reason is that optimality is not always preserved by adding or removing elements from a single optimal set. However as we discuss in sections 2.2 and 2.3 an enlarged class of optimal and near optimal sets has a structure that is preserved under the addition of elements and when the sets are large enough, the removal of elements as well. These properties can lay the groundwork for approximation and discovery of optimal sets.

### 2.2 Optimal and Near Optimal Sets

In section 1 a measure of the spread effectiveness of sets was introduced in (2). It will be convenient to convert this to a rank defined on subsets of $V$. In particular, suppose there exists a maximal match with $K$ vertices. We will order all non-empty subsets $A \subseteq V$ such that $|A| \leq K$ with a ranking function $\rho(A)$ defined as,

$$\rho(A) = \frac{F_{\max} - F(A)}{F_{\max} - F_{\min}}$$

where $F_{\max} = \max_{A \subseteq V, |A| \leq K} F(A)$, and $F_{\min}$ is the corresponding minimum. $F_{\min}$ can be calculated by computing $F$ for a maximal match of cardinality $K$, while $F_{\max}$ is the maximal value of $F$ among all one element subsets. We assume that $F_{\max} \neq F_{\min}$.

If $A$ is optimal then $\rho(A) = 1$ conversely the the worst performing set has value 0. For a constant $c$, $0 < c \leq 1$ and $K$, the non-empty set

$$L_{c,K} = \{ A : A \subseteq V, |A| \leq K, \rho(A) \geq c \}$$

4
defines a set of optimal and near optimal subsets, with the degree of near optimality depending of course on \( c \). Figure 12 shows an example of this set for the graph in EXAMPLE 2. It consists of a listing of the two element sets of nodes (column a) and three element sets of nodes (columns b-e) that are contained in \( L_{c,K} \) for \( c = 7/8 \) and \( K = 8 \). The sets of a fixed cardinality are listed according to rank as defined by \( \rho \).

The structure of optimal and near optimal sets is conveniently described in terms of a concept in combinatorial optimization known as a greedoid \[8, 2\].

**Definition 3** Let \( E \) be a set and let \( F \) be a collection of subsets of \( E \). The pair \((E, F)\) is called a **greedoid** if \( F \) satisfies

- \( G1 : \emptyset \in F \)
- \( G2 : \text{For } A \in F \text{ non-empty, there exists an } a \in A \text{ such that } A \setminus \{a\} \in F \)
- \( G3 : \text{Given } X, Y \in F \text{ with } |X| > |Y|, \text{ there exists an } x \in X, \text{ such that } Y \cup \{x\} \in F \)

A set in \( F \) is called feasible. Note that \( G2 \) implies that a single element can be removed from a feasible set \( S \) so that the reduced set is still feasible. By repeating this process eventually the empty set is reached. Conversely starting from the empty set, \( X \) can be built up in steps using the \( G3 \) property in a fashion resembling the greedy algorithm.

Turning to Figure 12 for \( L_{7/8,8} \), note that for any two element set \( Y \) and a three element set \( X \), there exists an element \( x \in X \setminus Y, \text{ such that } Y \cup x \in L_{c,K} \). For example when \( Y = \{3, 7\} \) in column (a) and \( X = \{2, 5, 8\} \) in column (c), we can choose the element \( 5 \in X \) to obtain \( \{3, 5, 7\} \) in column (b). In fact we will show that \( L_{c,K} \) satisfies condition \( G3 \) of the definition for \( 0 < c \leq 1, 0 \leq K \leq N \) (Proposition 1). The proof depends on several short lemmas. The first uses an adaptation of an argument in Clark et al.

**Lemma 1** Let \( S \subseteq V, u \in V \setminus S \). Then \( F(S) \geq F(S \cup \{u\}) \).

**Proof:** Suppose \( S \), a set of nodes is a target set for the random walk. Let \( E_{ij}^l(S) \) be the event, \( E_{ij}^l(S) = \{X_0 = i \in V, X_1 = j \in V \setminus S, X_r \notin S, 0 \leq r \leq l\} \). Thus paths of the random walk start at \( i \) and arrive at \( j \) without visiting \( S \) during the interval \([0, l]\). Also define the event \( F_{ij}^l(S, u) = E_{ij}^l(S) \cap \bigcup_{m=0}^{l} \{X(m) = u\} \). Paths in this event also start at \( i \) and arrive at \( j \) without visiting \( S \), but must visit the element \( u \) at some time during the interval \([0, l]\). Since a path either visits \( u \) in the time interval \([0, l]\) or it does not, it follows that:

\[
E_{ij}^l(S) = E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S, u) \tag{10}
\]

We have \( E_{ij}^l(S \cup \{u\}) \cap F_{ij}^l(S, u) = \emptyset \). This implies that,

\[
\chi(E_{ij}^l(S)) = \chi(E_{ij}^l(S \cup \{u\})) + \chi(F_{ij}^l(S, u)) \tag{11}
\]

and therefore:

\[
\chi(E_{ij}^l(S)) \geq \chi(E_{ij}^l(S \cup \{u\}) \tag{12}
\]

Here \( \chi(A) \) is the indicator function of the set \( A \). Recalling that \( T_S \) is the hitting time for set \( S \), the following relation comes from taking the expection of \( \chi(E_{ij}^l(S)) \) on the left hand side of \(12\) summing over all \( j \in V \setminus S \). Here \( \mathbb{E} \) denotes expectation.

\[
\text{Prob}\{T_S > l | X_0 = i\} = \mathbb{E} \left( \sum_{j \in V \setminus S} \chi(E_{ij}^l(S)) \right) \tag{13}
\]

5
A similar result is obtained for \( T_{S \cup \{u\}} \) from taking the expectation of \( \chi(E^i_j(S \cup \{u\})) \) on the right hand side of (12) and summing over \( j \in V \setminus S \). Summing once again over all \( l \) results in the inequality,

\[
h(i, S) \geq h(i, S \cup \{u\})
\]  

(14)

Finally on summing (14) over all \( i \) and recalling the definition of \( F \) (equation (2)) one obtains the result to be proved. \( \square \)

The following result uses that fact that \( F \) is supermodular.

**Lemma 2** For \( \bar{c} > 0 \), let \( \mathcal{P} = \{ X \subseteq V : F(X) \geq \bar{c} > 0 \} \). If \( A, B \in \mathcal{P} \) where \( |A| = |B| \) and \( |A \cap B| = |A| - 1 \), then \( A \cap B \in \mathcal{P} \).

**Proof:** The hypothesis implies the existence of a set \( X \) such that \( A = X \cup \{a\} \) and \( B = X \cup \{b\} \) with \( a \neq b, a, b \in V \). The supermodular property of \( F \) implies that:

\[
F(X \cup \{a\} \cup \{b\}) + F(X) \geq F(X \cup \{a\}) + F(X \cup \{b\})
\]  

(15)

Rearranging we have,

\[
F(X) \geq F(X \cup \{a\}) + [F(X \cup \{b\}) - F(X \cup \{a\} \cup \{b\})]
\]  

(16)

Thus on writing \( X = A \cap B \), using the hypothesis on \( A \), and then applying Lemma 1 to the bracketed quantity, we have,

\[
F(A \cap B) \geq F(A) \geq \bar{c}
\]

\( \square \)

The following lemma is part of a result in [3, 2] on paving greedoids.

**Lemma 3** If \( \mathcal{P} \) is any class of sets satisfying the conclusion of Lemma 1, \( \mathcal{S} = 2^V \setminus \mathcal{P} \) has property G3. That is, given any \( A, B \in \mathcal{S} \), with \( |A| > |B| \), there is an \( a \in A \setminus B \) such that \( B \cup \{a\} \in \mathcal{S} \).

**Proof:** Suppose the conclusion is false. If \( |A \setminus B| = 1 \), then for \( a \in A \setminus B \), \( B \cup \{a\} \in \mathcal{S} \). But \( A = B \cup \{a\} \), so this is a contradiction. Suppose next that \( |A \setminus B| > 1 \). Then there exists \( a \in A \setminus B \). We have \( B \cup \{a\} \) and \( B \cup \{a\} \in \mathcal{P} \). Thus the conclusion of Lemma 1 implies that \( B \in \mathcal{P} \), which is also a contradiction. \( \square \)

**Proposition 1** For \( 0 < c \leq 1 \) and \( 0 < K \leq N \), let \( L_{c,K} \) be the class of sets defined in equation (4). Then \( L_{c,K} \) satisfies condition G3.

**Proof:** If \( \mathcal{P} \) is the set defined in Lemma 1 then \( L_{c,K} = \mathcal{S} \) for some \( c \). In fact we may set \( c = \frac{F_{\text{max}} - \bar{c}}{F_{\text{min}} - F_{\text{min}}} \). If \( F_{\text{min}} \leq \bar{c} \leq F_{\text{max}} \), we have \( 0 \leq c \leq 1 \). \( L_{c,K} \) satisfies the conclusion of Lemma 3 and therefore it satisfies property G3. \( \square \)

Although most of the three element sets in Figure 12 also satisfy condition G2 a number of sets especially the lowest ranking ones fail to do so. In fact there are no one element sets in \( L_{c,K} \) so G2 does not hold for any two element set. Therefore let us suppose that \( c \) in (9) is such that \( c^*_m > c > c^*_{m-1} \), where \( c^*_m = \max_{|A| \\leq m} \rho(A) \). Then \( L_{c,K} \) only contains subsets of cardinality \( m \) or higher. In our example \( m = 2 \). Let \( G_m = \{ A : A \in L_{c,K}, |A| = m \} \) and \( R_{c,K} = \{ C : C \subseteq G_m \} \). A class of subsets of \( L_{c,K} \) and \( R_{c,K} \) is constructed in [5] that satisfies conditions G1–G3 and it is therefore a greedoid over \( V_K = \{ A \subseteq V : |A| \leq K \} \). Denoting the feasible sets of the greedoid by \( \mathcal{F}(L_{c,K}) \), it can be shown that feasible sets in \( \mathcal{F}(L_{c,K}) \cap L_{c,K} \) are elements of \( L_{c,K} \) and when the cardinality of a set exceeds \( m \) it satisfies G2 and G3 in such a way that the resulting set is still in \( \mathcal{F}(L_{c,K}) \cap L_{c,K} \). Here we state a theorem about the construction of a greedoid from \( L_{c,K} \):
Theorem 1 (see [5]) Let $L_{c,K}$ be the set defined in (9), where $c_m^* > c > c_{m-1}^*$ for some $m > 1$ and $0 < K \leq N$. There exists a class of subsets $\mathcal{F}(L_{c,K})$ of $L_{c,K} \cup R_{c,K} \cup \emptyset$, such that $(V_K, \mathcal{F}(L_{c,K}))$ is a greedoid.

This suggests that we can construct optimal or near optimal sets in a stepwise fashion starting from subsets of $G_m$ in $\mathcal{F}(L_{c,K})$. To do this, knowledge of $G_m$ is required and must be obtained by evaluating $F$ on subsets of $V$ up to cardinality $m$. The size of $m$ depends on the degree of optimality in the sense of (9). In general a larger the value of $c$, requires a larger value of $m$. We find in practice that $m$ is small relative to the number of nodes $N$, but further research is needed.

Suppose we seek an optimal set of cardinality $m < p < K$. Recall that in section 1, the greedy algorithm starts from the empty set and optimally adds elements of $V_K$ in a stepwise fashion until a set of cardinality $p$ is reached. On the other hand using the methods of the section 2.1 we could generate a maximal match of cardinality $K$ and look for best subsets of cardinality $p$. In the first case the degree of optimality is limited, and in the second, we cannot guarantee that the set of vertices of the match contain such an optimal subset. The optimal and near optimal sets in $\mathcal{F}(L_{c,K})$ of cardinality $m < p \leq K$ are the endpoint of a sequence of increasing sets from the empty set using property G3. In contrast to the greedy algorithm discussed in section 1, the initial sets of the construction are subsets of $G_m$ but are not necessarily optimal or near optimal sets. However once sets of cardinality greater than $m - 1$ are reached, feasible sets (i.e. sets in $\mathcal{F}(L_{c,K})$) have a degree of optimality $c$ defined in relation to all subsets up to cardinality $K$. If $c$ is large it is natural to expect that a feasible set of cardinality $p$ is highly ranked among all $p$ subsets when $c$ is large. In addition such a set contains elements of $L_{c,K}$ as subsets, so that any set of cardinality $p > m$ is optimal or near optimal and it contains optimal or near optimal subsets of cardinality $m$ or higher. If $X_k$ is the set constructed at the $k$th stage then $X_k \in G_k$ for $k \geq m$ and $X_k \subseteq X_{k+1}$. Thus the set $X_p$ i.e. some feasible set of cardinality $p$, is the approximation to the desired optimal set. To improve the approximation or to generate other solutions one could alternatively generate a set of vertices of a maximum match $\mathcal{M}$ (or vertex cover) of cardinality $K$. If it contains the vertices of a set in $G_p \cap \mathcal{F}(L_{c,K})$, then it is a feasible set and therefore has optimal and near optimal subsets of cardinality $p$ that can be compared to $X_p$. More generally, a feasible vertex cover will have subsets that are optimal or near optimal in the sense of (9). This is useful because such sets can be expected to be optimal or near optimal in the sense of problem (4).

Proposition 2 [5] Let $\mathcal{M}$, be a set of vertices from a maximal match (or vertices of a vertex cover). If $\mathcal{M}$ is a feasible set of the greedoid in Theorem 1, and has more than $m$ elements, then it contains optimal or near optimal feasible subsets in the sense of (9).

In the following section we take a first step towards systematizing the exploration of optimal and near optimal feasible sets by introducing moves among ”adjacent sets”.

2.3 The Graph of Optimal and Near Optimal Sets

To facilitate local search and comparison of optimal and near optimal feasible subsets we introduce a graph $\mathcal{G}(L_{c,K})$ whose nodes are the feasible sets of the greedoid identified in Theorem 1. To simplify the notation we use $\mathcal{F} = \mathcal{F}(L_{c,K})$ to denote the node set of $\mathcal{G}(L_{c,K})$. The local structure of the graph is defined by adjacent nodes.

Definition 4 Two nodes $A$ and $B \in \mathcal{F}$ are adjacent in $\mathcal{G}(L_{c,K})$ if one of the following statements is true.
• $A, B \in F$, $B = A \cup \{r\}$, for $r \notin A$

• $B = A \setminus a$ for some element $a \in A$

• $|A| = |B|$ and $|A \setminus B| = 1$

Given two feasible sets of equal cardinality say $A$ and $C$, in $F \cap L_{c,K}$, the set $D = A \setminus a$ is in $L_{c,K}$. By Theorem 1, such an $a$ always exists if $|A| > m$ (see section 2.2). We will assume this. Moreover there is a $d \in C$, not in $A$, such that $B = D \cup \{d\} \in F$ (by Proposition 1). The sets $A$ and $B$ and $D$ are adjacent in $G(F_{c,K})$ where clearly $B$ is the result of replacing $a$ by $d$ in $A$. By repeating this procedure one can construct a neighborhood of $A$ suitable for local search. When $A$ is a subset of vertices of a feasible maximal match or vertex cover, navigation to an enlarged neighborhood can be achieved by a sequence of moves to feasible adjacent sets. In order to reach feasible sets of size larger than $m$, some nodes corresponding to sets in $F$ of cardinality less than $m$ must first be determined by evaluating $F$ on these sets. A topic for future research is the development of efficient methods for doing this calculation as well as navigating $G(L_{c,K})$ so that the number of evaluations of $F$ is minimized.

3 Conclusion

We posed the problem of identifying the subset of nodes in a network that will enable the fastest spread of information in a decentralized communication environment. In a model of communication based on a random walk on an undirected graph $G = (V,E)$, the optimal set of nodes are found by minimizing the sum of the mean times of first arrival to the set by walkers who start at nodes outside the set.

Since the objective function for this problem is supermodular, the greedy algorithm has been a principal method for constructing approximations to optimal sets. References [3], [1] obtain results guaranteeing that these sets are in some sense within (1-1/e) of optimality. In this work we took a different approach. Rather than seeking an optimizing set for problem (4) without any information other than its cardinality and the graph $G$, we sought a solution within the class of optimal and near optimal sets $L_{c,K}$, where the degree of optimality $c$ is defined by ranking sets of cardinality up to $K$ according to their value of $F$ (see equation (9)).

In section 2, Theorem 1 was used to show the existence of a greedoid whose feasible sets include a subset of optimal and near optimal sets in $L_{c,K}$. In what follows we will call these feasible sets $F_{c,K}$. The supermodularity of $F$ can be used to show that adding a single element to $F_{c,K}$ produces another set in $F_{c,K}$ and if the original set is larger than $m$, the minimum size of sets in $F_{c,K}$, there is some element which can be removed so that the resulting set is still in $F_{c,K}$ (see [5] for details). A significant consequence of Theorem 1 as stated in Proposition 2 is that a feasible vertex cover larger than $m$ contains optimal and non-optimal subsets in the sense of [9] so the search for solutions and approximations of the problem (4) can begin there. Every feasible set can be constructed element by element beginning with elements contained in sets of $F_{c,K}$ of cardinality $m$. When a feasible set of cardinality $p < K$ in $L_{c,K}$ is an initial approximation to the solution of equation (4) for $p = M$, then a local search of adjacent sets (see Definition 2.4) using the addition and removal of elements can be done to look for improvements. These local moves are conveniently represented in terms of navigation through a graph whose nodes are the feasible sets of the greedoid discussed in Theorem 2.1.

In conclusion, the framework just presented allows the problem of optimal spread to be considered in the context of the topology of the graph of optimal and near optimal sets—the greedoid graph (see section 2.3). In particular, one can investigate subsets of feasible maximal matches (or
other vertex covers) as candidate optimal sets. Given $M$, $c$ and $K$, how can these sets be efficiently computed? Understanding the relationship between the topology of the original graph $G$ and the greedoid graph is an important next step to answering this question and is a topic of future research. It is also of interest to investigate the extent to which this framework applies to other modes of distributed communication besides irreducible Markov chains and random walks.

Acknowledgement: The author thanks Jason Wu of the University of California, Berkeley for his assistance in the initial stages of this project.

References

[1] V.S. Borkar, J. Nair, N. Sanketh, Manufacturing Consent, 48th Annual Allerton Conference, Allerton House, UIUC, Illinois, September 2010, pp. 1550-1555

[2] A. Bjorner, G. Ziegler, Introduction to Greedoids, in Matroid Applications (ed. N. White), Encyclopedia of Mathematics, Vol. 40, Cambridge University Press, London, UK, 1992, pp.284-357

[3] A. Clark, L. Bushnell, R. Poovendran, Leader Selection for Minimizing Convergence Error in Leader-Follower Systems: A Supermodular Optimization Approach, 10th International Symposium Modeling and Optimization in Mobile, Ad-Hoc and Wireless Networks (WiOpt), May 2012, pp. 111-115

[4] T.H. Cormen, C.E. Leiserson, R.L. Rivest, C. Stein Introduction to Algorithms, third edition, MIT Press, 2009

[5] F. Hunt, The Structure of Optimal and Near Optimal Sets in Consensus Models, preprint 2014

[6] D. Jungnickel, Graphs, Networks, and Algorithms, Springer Verlag, New York, Berlin, Heidelberg, Tokyo, 1991

[7] J. Kemeny, J. Snell, Finite Markov Chains, 2nd edition, Springer-Verlag, New York, Berlin, Heidelberg, Tokyo, 1976

[8] B. Korte, L. Lovasz, R. Schrader, Greedoids, Algorithms and Combinatorics Series, Vol. 4, Springer Verlag, Berlin, Germany, 1991

[9] R. Lambiotte, R. Sinatra, J.C.Delvenne, T.S. Evans, M. Barahona, V.Lattora, Interweaving dynamics and structure, Phys Rev. E 84, 017102, 2011

[10] G.L. Nemhauser, L.A. Wolsey, M.L. Fisher, An analysis of approximations for maximizing submodular set functions-I, Mathematical Programming, Vol. 14, pp.553-574, 2003
Figure 1: graph in EXAMPLE 1 showing $K=1$ optimal set

Figure 2: graph in EXAMPLE 1 showing $K=2$ optimal set

Figure 3: graph in EXAMPLE 1 showing $K=3$ optimal set
Figure 4: graph in EXAMPLE 1 showing K=4 optimal set

Figure 5: graph for EXAMPLE 2 with 9 vertices showing optimal K=1 set
Figure 6: graph for EXAMPLE 2 showing K=2 optimal set

Figure 7: graph for EXAMPLE 2 showing optimal set for K=3
Figure 8: graph for EXAMPLE showing optimal set K=4

Figure 9: graph for EXAMPLE 2 showing optimal set for K=5
Figure 10: graph for EXAMPLE 2 with optimal set $\mathcal{M} = \{1, 3, 5, 6, 7, 8\}$ for $K = 6$. $\mathcal{M}$ is also a vertex cover. It has subsets that are optimal sets for $K = 1, 2, 3, 4$.

Figure 11: graph for EXAMPLE 3 is shown with two vertex covers. The nodes of the covers are colored. On the left, the vertex cover contains no optimal subsets except itself, the vertex cover shown on the right contains optimal subsets for $K = 1$ through 4.
Figure 12: Two (see column a) and three element sets (columns b-e) ranked by $\rho$ for optimal and near optimal sets of the graph in Figure 10. The sets are contained in $L_{c,K}$ for $c = 7/8$, $K = 8$, equation (9).

|       | (a) | (b) | (c) | (d) | (e) |
|-------|-----|-----|-----|-----|-----|
| 3 7   | 3 4 | 9   | 1 5 | 8   | 1 7 |
| 1 7   | 3 6 | 8   | 1 5 | 9   | 1 7 |
| 2 7   | 1 5 | 7   | 2 5 | 8   | 2 7 |
| 3 8   | 2 5 | 7   | 2 5 | 9   | 2 7 |
| 3 9   | 3 4 | 7   | 3 8 | 9   | 1 4 |
| 3 4   | 3 5 | 7   | 3 5 | 8   | 1 6 |
| 3 6   | 3 7 | 9   | 3 5 | 9   | 2 4 |
| 1 8   | 3 6 | 7   | 1 6 | 7   | 2 6 |
| 2 8   | 3 7 | 8   | 2 6 | 7   | 1 8 |
| 1 9   | 3 4 | 6   | 1 4 | 7   | 2 8 |
| 2 9   | 1 4 | 6   | 2 4 | 7   | 1 2 |
| 5 7   | 2 4 | 6   | 3 4 | 8   | 1 4 |
| 1 4   | 1 4 | 9   | 3 6 | 9   | 1 5 |
| 1 6   | 1 6 | 8   | 1 3 | 7   | 2 4 |
| 2 4   | 2 4 | 9   | 2 3 | 7   | 2 5 |
| 2 6   | 2 6 | 8   | 1 7 | 8   | 1 3 |
|       |     |     |     |     | 5 8 |

The sets are contained in $L_{c,K}$ for $c = 7/8$, $K = 8$, equation (9).
[11] C. Papadimitriou, D. Steiglitz, *Combinatorial Optimization and Complexity*, Chapter 17 and p.432, Dover Publications, 1998