Volume 121 (2016) http://dx.doi.org/10.6028/jres.121.008
Journal of Research of the National Institute of Standards and Technology

# *An Algorithm for Identifying Optimal Spreaders in a Random Walk Model of Network Communication*

**Fern Y. Hunt**

National Institute of Standards and Technology,
Gaithersburg, MD 20899

fern.hunt@nist.gov

In a model of network communication based on a random walk in an undirected graph, what subset of nodes (subject to constraints on the set size), enables the fastest spread of information? The dynamics of spread is described by a process dual to the movement from informed to uninformed nodes. In this setting, an optimal set *A* minimizes the sum of the expected first hitting times *F*(*A*), of random walks that start at nodes outside the set. Identifying such a set is a problem in combinatorial optimization that is probably NP hard. *F* has been shown to be a supermodular and non-increasing set function and fortunately some results on optimization of such functions exist, e.g., in the work of Ilev.

In this paper, the problem is reformulated so that the search for solutions to the problem is restricted to a class of optimal and "near" optimal subsets of the graph. We introduce a submodular, non-decreasing rank function $\rho$, that permits some comparison between the solution obtained by the classical greedy algorithm and one obtained by our methods. The supermodularity and non-increasing properties of *F* are used to show that the rank of our solution is at least $(1-\frac{1}{e})$ times the rank of the optimal set. When the solution has a higher rank than the greedy solution this. constant can be improved to $(1+\chi)(1-\frac{1}{e})$ where $\chi > 0$ is determined a posteriori. The method requires the evaluation of *F* for sets of some fixed cardinality *m*, where *m* is much smaller than the cardinality of the optimal set. Given $\nu = \rho(A)$ , a class of examples is presented that illustrate the tradeoff between *m* and solution quality $\nu$ .

**Key words:** consensus models; first hitting time; greedoids; networks; random walk; submodular; supermodular functions.

**Accepted:** March 25, 2016

**Published:** May 10, 2016

http://dx.doi.org/10.6028/jres.121.008

## 1. Introduction

The study of information spread (or dually consensus) in complex networks has been the subject of intense research in the past decade [1-5] where the role of distinguished subsets of nodes, such as "leaders" in consensus models and "influential spreaders" in models of information spread, is studied. In particular the research reported in references [3-5] developed methods for obtaining optimal spreaders –as determined by some measure of subset performance. A substantial body of related work is concerned with the construction and performance analysis of algorithms for efficient information spread for example the so-called push/pull algorithms [6], the independent cascade model [5], a random averaging scheme [7] and the GOSSIP model of Ref. [8]. In this paper, our focus will be on the identification of optimal spreaders in a network. We will use a random walk communication model and an objective function associated with this process. Results of this research are relevant to the design of algorithms for routing in wireless communication systems when location information is not available [2, 9], identification of influential individuals in a social network [3] and in sensor placements for efficiently detecting intrusions in water networks [10].

Given a connected graph $G=(V,E)$ with $N$ vertices $V$ and edges $E$, information spreads through the network by a process that is dual to the direction of the random walk (see Ref. [11]). An optimal spreader in our setting is defined in terms of a set function $F$ where for a subset $A \subset V$, $F(A)$ is the sum of mean first arrival times to $A$ by random walkers that start at nodes outside of $A$. If $A$ is an effective target set for the random walks (dually an effective spreader) then $F(A)$ is small. Thus, the optimal set (subject to a cardinality constraint $K$) minimizes $F(A)$ subject to $|A| \le K$,

$$\min_{A \subset V,\, |A| \le K} F(A). \tag{1}$$

Recall that a random walker situated at a node $i \in V$, moves to a neighboring node $j \in V$ in a single discrete time step with probability,

$$Prob\{i \to j\} = \{p(i,j) > 0, \quad if \quad (i,j) \in E, \quad p(i,j) = 0, \quad if \quad (i,j) \notin E\} \tag{2}$$

**NOTE:** In this paper $p(i,j) = 1/deg(i)$ where $deg(i)$ is the degree of node $i$. However, any probabilities for which the resulting Markov chain is ergodic can be used.

The matrix $\mathbb{P} = (p_{ij})_{i,j=1\cdots N}$ is the transition matrix of a Markov chain which in our choice or any choice of transition probabilities, is assumed to be irreducible and aperiodic [12]. Starting at any node outside of $A$, 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 expected hitting time is $\mathbb{E}[T_A]$. If the walker starts at a fixed $i \notin A$, then the expected hitting time is the conditional expectation $\mathbb{E}[T_A \mid X_0 = i] = \mathbb{E}_i[T_A]$. Writing $h(i,A) = \mathbb{E}_i[T_A]$, the value of $F$ at $A$ is expressed as

$$F(A) = \sum\nolimits_{i \notin A} h(i,A). \tag{3}$$

Given $A$, $F(A)$ can be evaluated by solving a suitable linear equation. Indeed a standard result in Markov chain theory [12] tells us that $h(i,A)$ is the ith component of the vector $\mathrm{H}$, which is the solution of the linear equation,

$$\mathrm{H} = \mathbf{1} + \mathbb{P}_A \mathrm{H}, \tag{4}$$

where $\mathbf{1}$ is a column vector of $N - |A|$ ones and $\mathbb{P}_A$ is the matrix that results from crossing out the rows and columns of $\mathbb{P}$ corresponding to the nodes of $A$. The value $F(A)$ is then the sum of the components of $\mathrm{H}$.

Borkar, Nair and Sanketh, [13] showed that for subsets $A \subseteq B \subseteq V$ and $j \in V$, $F(A) - F(A \cup \{j\}) \ge F(B) - F(B \cup \{j\})$, that is, $F$ is a supermodular function. Thus $-F$ is submodular so if it is bounded our problem is an instance of submodular maximization, a classic problem in combinatorial optimization. In 1987, Nemhauser, Wolsey and Fisher [14] showed for a bounded submodular function that a set constructed by the greedy algorithm has an approximation ratio of $(1-1/e)$. More recently, Borgs, Brautbar, Chayes and Lucier [5] and Sviridenko, Vondrak, and Ward [15], showed that approximations of comparable quality could be obtained very efficiently using different methods. To minimize the convergence rate to consensus of a leader-follower network, Clark, Bushnell and Poovendran [16] considered a supermodular function closely related to ours and showed that the greedy algorithm produces an approximation that is within $(1-1/e)$ of optimal.

In this paper we will discuss a method that obtains an exact or approximate solution to Eq. (1) by introducing additional constraints in the problem. These constraints are based on properties of the underlying graph. Recall that a vertex cover is a set of vertices that cover every edge of the graph. That is, every edge of the graph is incident to vertex in this set. Observing that a vertex cover of the graph with $C$ vertices is an optimal set for $K = C$, sets of cardinality $C$ or less can be assigned a ranking relative to it. Using the rank (introduced in Sec. 2.2), we define a class of optimal and near optimal sets $L_{\nu,C}$, where $\nu$ is the minimum rank of sets in the class. Here we consider $\nu$ as a measure of the quality of the approximation. To solve the problem for $K < C$, we choose a collection of sets $\mathbf{S} \subset L_{\nu,C}$. Each set in $\mathbf{S}$ has cardinality $m$ where $m$ is the minimum cardinality of sets in $L_{\nu,C}$. Note that the exact solution is in $L_{\nu,C}$ for $m < K \le C$. The output of this method is the best set that results from a greedy extension of each set in $\mathbf{S}$, to a set of cardinality $K$. The method requires the determination of sets of cardinality $m$ each of pre-determined quality $\nu$ and the computational effort involved as discussed in Sec. 4 is $O(N^{m+3})$. We assume that $m \ll K$ so a natural question is given $m$ what quality $\nu$ can be expected? Conversely given a required solution quality $\nu$, what $m$ is needed?

The plan of the paper is as follows: Sec. 1 contains a definition and discussion of optimal and near optimal sets ranked relative to a vertex cover of the graph $G$ of cardinality $C$. In Sec. 2.2 we demonstrate how the method is applied to a graph using a collection of sets $\mathbf{S}$ that are subsets of the vertex cover. If every vertex cover contained optimal sets as subsets, it would make sense to use this choice consistently. Unfortunately, optimality of a set is generally not preserved by the addition or deletion of elements, otherwise the greedy algorithm would always yield exact solutions. We remedy this situation in part by selecting a group $\mathbf{S}$ of $m$ element sets in $L_{\nu,,C}$ that contain a class of subsets satisfying certain properties (see Sec. 3) for all sets up to cardinality $m$. By adjoining all supersets of $\mathbf{S}$, to $\mathbf{S}$ itself and these subsets, one obtains a greedoid [17] as discussed in Sec. 3. Its feasible sets are closed under the addition and deletion of certain elements. Moreover all feasible sets of cardinality $n > m$ are in $L_{\nu,C}$ and are therefore optimal or near optimal. In general, the greedoid is not unique and it may or may not contain optimal sets of required cardinality $K$. However, any offered solution of our method that is feasible will be near optimal with some pre-defined quality. Sufficient conditions for the existence of $\mathbf{S}$ are stated in Sec. 3 but the details of the greedoid construction can be found in [18].

In Sec. 3 we demonstrate the method on a second graph where $\mathbf{S}$ is chosen to be a group of feasible sets of a greedoid. Since we do not as yet have sufficient conditions that guarantee the optimal solution is a greedy extension of $\mathbf{S}$, a complete proof of the effectiveness of this approach will require further research. In Sec. 4, the quality of the approximation is evaluated in terms of the ranking function $\bar{\rho}$ introduced in Sec. 2.2. After normalizing $F$, we obtain $\rho$, a bounded submodular set function with $\rho(\emptyset) = 0$. We can apply the results in Ref. [14], to show that the ratio of the rank of our approximation to that of the optimal set is at least $(1-\frac{1}{e})$. Moreover, the approximation can be compared to the other solutions obtained by the greedy extension of sets of cardinality less than $m$ including the classic greedy method that starts with a one element set. In particular, if the rank of a greedy solution is less than $\nu$, or if the greedy solution at stage $m$ is in $\mathbf{S}$, then the solution $S^*$ obtained by our method satisfies an inequality that improves the $(1-\frac{1}{e})$ bound,

$$\rho(S^*) \ge (1+\chi)(1-\frac{1}{e})\rho(\mathcal{O}_K) \tag{5}$$

where $\chi > 0$ is a constant determined *a posteriori* and $\mathcal{O}_K$ is an optimal solution of Eq. (1). Finally in Sec. 5, a probabilistic representation of the increment $F(S \cup \{u\}) - F(S)$ is given in terms of first hitting times of the respective sets. It is not directly used in the rest of the paper but it is of independent interest.

## 2. Finding and Approximating Optimal Sets

### 2.1 Maximal Matches

The optimization problem as posed in Eq. (1) assumes no advance knowledge about the optimal set or any other possibly related sets. We first consider a process of obtaining optimal sets by using subsets of existing ones. Let *A* be a vertex cover (not necessarily a minimal one). 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 Eq. (4) implies that $h(i, A) \geq 1$ so it follows that *A* must be an optimal set for its own cardinality. Thus a solution for $C = |A|$ is obtained by constructing a vertex cover. Fortunately, a maximal match can be constructed by a simple greedy algorithm and its vertices are a vertex cover with cardinality $C \leq 2*(\mathcal{VC})$ where $\mathcal{VC}$ is the cardinality of a minimum vertex cover [19]. Therefore, without loss of generality we turn our attention to the solution of Eq. (1) for $K \leq C$.

### 2.2 Optimal and Near Optimal Sets

We introduced a measure of the spread effectiveness of sets in Eq. (3). It will be convenient to convert this to a rank defined on subsets of $V$. In particular, suppose there exists a vertex cover with $C$ vertices. We will order all non-empty subsets $A \subseteq V$ such that $|A| \leq C$ with a ranking function $\bar{\rho}(A)$ defined as,

$$\bar{\rho}(A) = \frac{F_{max} - F(A)}{F_{max} - F_{min}} \quad (6)$$

where $F_{max} = \max_{\emptyset \neq A \subseteq V, |A| \leq C} F(A)$, and $F_{min}$ is the corresponding minimum. $F_{min}$ can be calculated by computing $F$ for a vertex cover of cardinality $C$ whose elements we assume are the endpoints of a maximal match. We define $F_{max}$ to be the maximum value of $F$ among all one element subsets. We assume that $F_{max} \neq F_{min}$. If this were not the case, $F(A)$ would be have the same value for any non-empty subset *A* with $|A| \leq C$. Thus any $A$ with $|A| \leq K$ would be a solution of the problem.

If *A* is optimal and $|A| = C$ then $\bar{\rho}(A) = 1$. Conversely, the worst performing set has value $0$. For a constant $\nu (0 < \nu \leq 1)$ and $C$, the non-empty set

$$L_{\nu,C} = \{A : A \subseteq V, |A| \leq C, \bar{\rho}(A) \geq \nu\} \quad (7)$$

defines a set of optimal and near optimal subsets, with the degree of near optimality depending on $\nu$. Let $m$ be the smallest cardinality of sets in $L_{\nu,C}$. Starting with a collection of sets $\mathbf{S} \subset L_{\nu,C}$ of size $m$, our method is to seek a solution to the optimization problem as posed in Eq. (1) by greedily augmenting each set until it reaches the desired size $K$. The offered approximation is the best (has the lowest $F$ value) of these extended sets. We can always find a $\nu$ and $C$ so that $L_{\nu,C}$ contains the optimal set of cardinality $K$, but we do not have a proof that the approximation generated by subsets of a vertex cover is optimal. However, since our solution is a superset of sets in $L_{\nu,C}$, it is also in $L_{\nu,C}$ and, therefore, has minimum rank $\nu$. We illustrate the method with an example.

Figures 2 through 6 show a graph with $N = 9$ vertices along with the vertices of optimal sets for $K = 1$ through $5$. To solve the problem for $K = 4$, we note that the class of optimal and near optimal sets based on $C = 8$ and $\nu = 0.90$ has minimum set size $m = 2$. The set $\mathcal{M} = \{1, 3, 5, 6, 7, 8\}$ is a vertex cover (calculated from the maximal match algorithm).We define $\mathbf{S}$ to be the two element subsets of $\mathcal{M}$ that are in $L_{.90,8}$. The first column of Fig. 1 lists these sets and subsequent columns show the results of one element extensions of $\mathbf{S}$ until $K = 5$. Optimal sets are shown in red. In this example the offered approximation is optimal. This is also the case for extensions up to $K = 5$. In this case we see that the method identifies optimal sets that are subsets of $\mathcal{M}$ as well as others that are not, e.g., $\{2,3,4,6,8\}$, underlining the fact that the method finds sets that are reachable by greedy extension of subsets of $\mathcal{M}$. The offered approximation for this method is guaranteed to be in $L_{.90,8}$. This is a consequence of Proposition 1 which is discussed and proved in Sec. 3.

| S | | | | | | |
|---|---|---|---|---|---|---|
| {3,7} | ⟶ | {3,5,7} | ⟶ | {3,5,7,8} | ⟶ | {1,3,5,7,8} |
| {1,7} | ⟶ | { 1,5,7} | ⟶ | {1,5,7,8} | ⟶ | {1,3,5,7,8} |
| {3,8} | ⟶ | {3,6,8} | ⟶ | {3,4,6,8} | ⟶ | {2,3,4,6,8} |
| {3,6} | ⟶ | {3,6,8} | ⟶ | {3,4,6,8} | ⟶ | {1 ,3,4,6,8} |

**Fig. 1.** Optimal sets of the graph in Fig. 2 for K=4 and 5 obtained by greedy extension of **S**, subsets of a vertex cover.

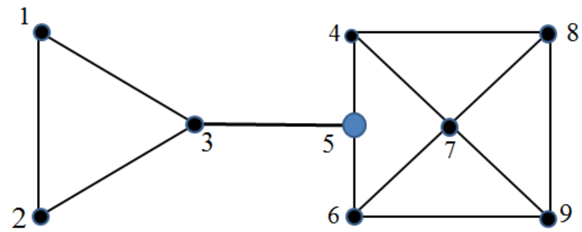


**Fig. 2.** Graph with N=9 vertices, showing optimal set for K=1.

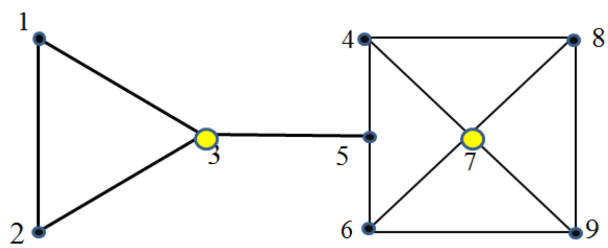


**Fig. 3.** Optimal set K=2 for graph in Fig. 2.

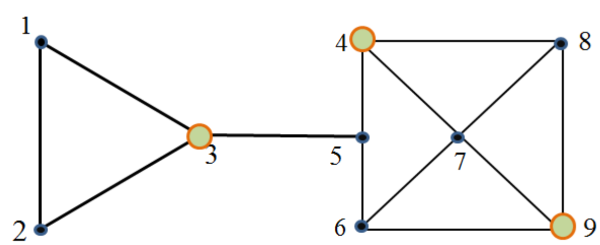


**Fig. 4.** Optimal set K=3 for graph in Fig. 2. The set {3,6,8} is also optimal by symmetry.

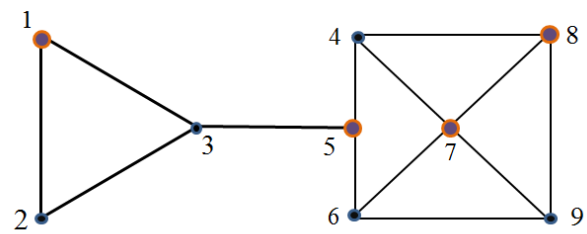


**Fig. 5.** Optimal set K=4 for graph in Fig. 2.

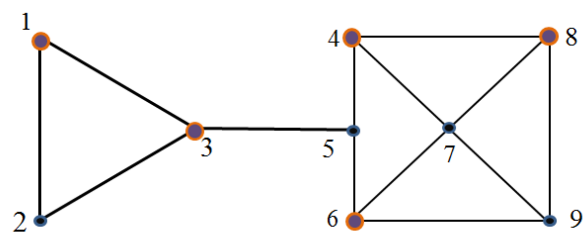


**Fig. 6.** Optimal set K=5 for graph in Fig. 2.

## 3. Closure Property of Optimal and Near Optimal Sets

In Sec. 2.2, we demonstrated our method for approximating a solution of optimization problem in Eq. (1) based on greedy extensions of subsets of a vertex cover that are optimal or near optimal. Unfortunately, a vertex cover can fail to have such subsets other than the vertex cover itself (see an example in Ref. [18]). This is the motivation for finding other classes of optimal and near optimal sets that permit the addition and deletion of elements. We conjecture that greedy extension of such sets will have the largest likelihood of success. The structure we seek is conveniently described in terms of a generalization of the matroid, known as a *greedoid* [17, 20].

**Definition 1** *Let* $\mathbf{E}$ *be a set and let* $\mathcal{F}$ *be a collection of subsets of* $\mathbf{E}$. *The pair* $(\mathbf{E},\mathcal{F})$ *is called a* *greedoid* *if* $\mathcal{F}$ *satisfies*

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

A set in $\mathcal{F}$ is called feasible. Note that **G2** implies that a single element can be removed from a feasible set $X$ so that the reduced set is still feasible. By repeating this process, the empty set eventually is reached. Conversely starting from the empty set, $X$ can be built up in steps through repeated use of **G3**.

We now show that $L_{c,K}$ satisfies condition **G3** of the definition for any $0 < c \le 1$, $0 \le K \le N$ (Proposition 1). The proof depends on the following lemma and uses an adaptation of an argument in Clark et al. [16].

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

**Proof:** Suppose $S$ -a set of nodes, is a target set for a random walk. Let $E_{ij}^{l}(S)$ be the event, $E_{ij}^{l}(S) =$ $\{X_0 = i \in V, X_l = j \in V \setminus S, X_r \notin S, 0 \le r \le l\}$. Thus, paths of the random walk in this event start at $i$ and arrive at $j$ without visiting $S$ during the interval $[0,l]$. Also define the event $F_{ij}^{l}(S,u) = E_{ij}^{l}(S) \cap \bigcup_{m=0}^{l} \{X(m) = u\}$ where $u \notin S$. 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:

$$E_{ij}^{l}(S) = E_{ij}^{l}(S \cup \{u\}) \cup F_{ij}^{l}(S,u) \tag{8}$$

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

$$\mathbf{1}_{E_{ij}^{l}(S)} = \mathbf{1}_{E_{ij}^{l}(S \cup \{u\})} + \mathbf{1}_{F_{ij}^{l}(S,u)} \tag{9}$$

and therefore:

$$\mathbf{1}_{E_{ij}^{l}(S)} \geq \mathbf{1}_{E_{ij}^{l}(S \cup \{u\}} \tag{10}$$

Here $\mathbf{1}_A$ is the usual indicator function of the set *A*, i.e., the function $\mathbf{1}_A : \Omega \to \{0,1\}$. Recalling that $T_S$ is the hitting time for set $S$, the following relation comes from taking the expectation of $\mathbf{1}_{E_{ij}^{l}(S)}$ on the left hand side of Eq. (10) summing over all $j \in V \setminus S$. Here $\mathbb{E}$ denotes expectation.

$$\mathbf{Prob}\{T_S > l \mid X_0 = i\} = \mathbb{E}\left( \sum_{j \in V \setminus S} \mathbf{1}_{E_{ij}^{l}(S)} \right) \tag{11}$$

A similar result is obtained for $T_{S \cup \{u\}}$ from taking the expectation of $\mathbf{1}_{E_{ij}^{l}(S \cup \{u\})}$ on the right hand side of Eq. (10) and summing over $j \in V \setminus S$. Summing once again over all $l \geq 1$ results in the inequality,

$$h(i,S) \geq h(i, S \cup \{u\}) \tag{12}$$

**Proposition 1** *For* $0 < c \leq 1$ *and* $0 < K \leq N$, *let* $L_{c,K}$ *be the class of sets defined in Eq. (7). Then* $L_{c,K}$ *satisfies condition* **G3**.

**Proof:** The conclusion follows from the definition of $L_{c,K}$ and Lemma 1. □

The proposition establishes that $L_{c,K}$ satisfies the **G3** property for greedoids. However, **G2** does not hold. For example if the set *A* has cardinality $m$ where $m$ is the size of the smallest set in $L_{c,K}$ then $A \setminus \{a\}$ cannot be in $L_{c,K}$ for any element $a \in A$. Conversely, let $c_n = \max_{|X| \leq n} \rho(X)$. If $c_m \geq c > c_{m-1}$ then $m$ is the size of the smallest set in $L_{c,K}$. Define $G_n$ to be all sets in $L_{c,K}$ of cardinality $n$. To create a class of sets with the **G2** property, one constructs subsets of $G_m$ of size $n \leq m$ that are "augmentable," i.e., that satisfy **G3**. Sets $G_n$ for $n > m$ are culled so the remaining sets are supersets of the "augmentable" sets and, therefore, satisfy **G2**. The greedoid will then consist of selected subsets and supersets of $G_m$. Conditions for the existence of "augmentable" subsets of $G_m$ and proof of the validity of the resulting greedoid construction can be found in Ref. [18]. Rather than repeat the details of these arguments here, we close this section with an example showing the greedoid of a graph (Fig. 7) and its use in the solution of Eq. (1).The minimum cardinality of a set in the class of optimal and near optimal sets $L_{.85,7}$ is $m = 3$.

These sets are used to create the greedoid depicted in Fig. 8. Note that **G1** - **G3** are satisfied. Assume the optimal set for $K = 4$ is unknown. Then our method in this case is to take **S** to be the three element sets in $L_{.85,7}$ that are feasible sets of the greedoid and perform a greedy extension of each set. In Fig. 8, a line is drawn between a set and its greedy extension. We have also drawn greedy extensions of sets of cardinality $n < m$ as well. The optimal sets are shown in red and so they are in the greedoid. The offered approximations are in fact exact.

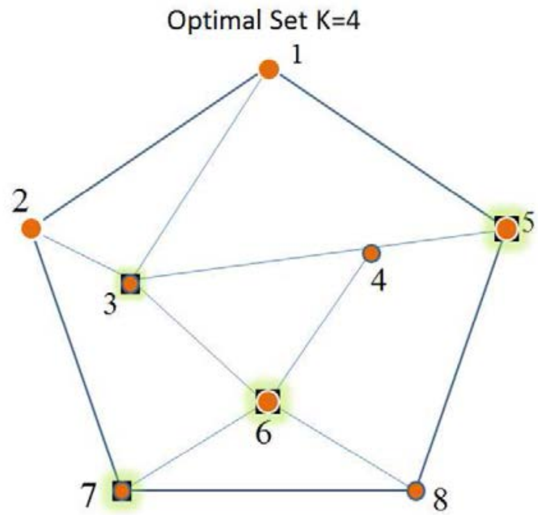


**Fig. 7.** Graph with N=8, vertices. Vertices of optimal set K=4 shown as squares.

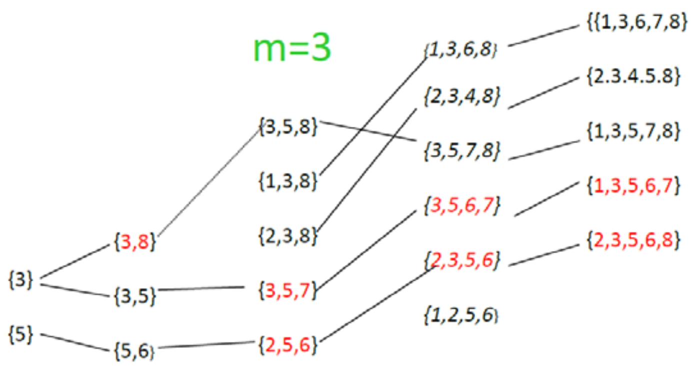


**Fig. 8.** Greedoid constructed from optimal and near optimal sets $L_{.85,7}$ of graph in Fig. 7. Empty set not shown.

## 4. Quality of the Approximation

### 4.1 Comparison with the Optimal Solution and Greedy Solution

Following Ilev [21], $F$ can be defined for the empty set as,

$$0 \le F(\emptyset) = \max_{X \cap Y = \emptyset, X,Y \subseteq V} F(X) + F(Y) - F(X \cup Y) < \infty \tag{13}$$

Thus, by the definition of $\bar{\rho}$, $\bar{\rho}(\emptyset) = \frac{F_{max} - F(\emptyset)}{F_{max} - F_{min}}$. This means the normalized function defined on sets $A$, $\rho(A) = \bar{\rho}(A) - \bar{\rho}(\emptyset)$ is bounded, submodular, non-decreasing. For the empty set we have $\rho(\emptyset) = 0$.

Our offered solution is the result of a greedy extension of a group of $m$ element sets $\mathbf{S}$. Using $\rho$ it can be compared to an $m$ element set that is the result of greedily adding single elements $m$ times. Call this set $S_g$. We first suppose that $S_g \in \mathbf{S}$.

**Lemma 2** *Suppose* $S_g \in \mathbf{S} \subseteq L_{\nu,C}$. *Let* $S_g^{(K)}$ *be the K element set obtained from the greedy extension. If* $S^*$ *is the offered solution, then*

$$F(S^*) \le F(S_g^{(K)}) \tag{14}$$

**Proof:** $F(S^*)$ is the minimum value of all the values obtained by the greedy $K - m$ extension of elements in $\mathbf{S}$. □

The set $S_g^{(K)}$ is also the result of greedily adding single elements $K$ times. Thus, we may use Sec. 4 in Ref. [14] and the definition of $\rho$ to conclude that:

**Corollary 1** *If* $S^*$ *is the solution constructed by the method described in Secs. 2.2 and 3 above, then*

$$\rho(S^*) \ge (1 - \frac{1}{e})\rho(\mathcal{O}_K^*) \tag{15}$$

where $\mathcal{O}_K^*$ *is the optimal solution of optimization problem in Eq. (1).*

Once $F(S^*)$ and $F(S_g^{(K)})$ have been computed we can determine $\chi$ such that $\rho(S^*) = (1+\chi)\rho(S_g^{(K)})$. Therefore, if $F(S^*) < F(S_g^{(K)})$ then the bound in Corollary 1 can be strengthened.

**Proposition 2** *When* $F(S^*) < F(S_g^{(K)})$, *so* $\chi > 0$, *then*

$$\rho(S^*) \ge (1+\chi)(1 - \frac{1}{e})\rho(\mathcal{O}_K^*) \tag{16}$$

If $S_g \notin \mathbf{S}$, the conclusion of Proposition 2 is still valid when the greedy extension to a $K$ element set $S_g^{(K)}$ that satisfies $\rho(S_g^{(K)}) < \nu$. Indeed by the closure property of $L_{\nu,C}$ (Proposition 1), $\rho(S^*) \ge \nu$ and thus $\chi > 0$. This bound is also valid for solutions obtained using the greedy extension of sets of cardinality less than $m$ for which lower bounds of the type $(1 - \frac{1}{e})$ have been established. A lower bound of $(1 - \frac{1}{e})$ was established by Borkar et al. [13]. Specifically, it is a lower bound on the ratio of $F(S_g) - F(\{a\})$ to $F(\mathcal{O}_k^*) - F(\{a\})$, where $S_g$ is the result of the greedy algorithm starting with singleton $a$.

### 4.2 Computational Effort and Tradeoff with Quality

A rough estimate of the complexity of the method follows from realizing that the collection $\mathbf{S} \in L_{\nu,C}$, has at most $\binom{N}{m}$, $m$ element sets. To determine whether or not a particular set is near optimal, Eq. (4) must be solved and this involves $O(N^3)$ operations. Thus $\mathbf{S}$ is determined in $O(N^{m+3})$ operations. The greedy extension of an $m$ element to a $K$ element set involves $O((K-m)(N-m)) = O(N^2)$ so that the extension of every set in $\mathbf{S}$ involves $O(N^{m+2})$ operations. Overall then, the method requires $O(N^{m+3}) + O(N^{m+2}) = O(N^{m+3})$ operations. Thus, it is desirable to make $m$ as small as possible. In fact we assume $m \ll K$. However, the size of $m$ affects the accuracy. Taking $\nu$ to be a measure of the quality of the approximation, we want to know given $m$ ,what $\nu$ can be expected? Conversely given a desired quality $\nu$, what $m$ is required? We will employ the elemental curvature of the rank function (see Eq. (6)). Elemental curvature was used by Wang, Moran, Wang, and Pan [22] in their treatment of the problem of maximizing a monotone non-decreasing submodular function subject to a matroid constraint. Recall from Sec. 4.1, that $\rho$ is a submodular, monotone and non-decreasing set function that vanishes on the empty set.

The elemental curvature of $\rho$ is defined over $L_{\nu,C}$ in terms of the marginal increase in the rank of a set when a single element is added to it. First let $A$ be a set and $i \notin A$,

$$\rho_i(A) = \rho(A \cup i) - \rho(A) \tag{17}$$

and then for a fixed $A \in L_{\nu,C}$ set,

$$k_{ij}(A) = \frac{\rho_i(A \cup j)}{\rho_i(A)}. \tag{18}$$

The curvature is defined then as,

$$\kappa = \max\{k_{ij}(A) : A \subset L_{\nu,C}, i \neq j, i, j \notin A\}. \tag{19}$$

Since $\rho$ is submodular $\kappa \leq 1$. This can be also be deduced from the increment formula (Eq. (25)). Now suppose $S \subset T \subset L_{\nu,C}$. Given $\nu$, we want to determine the minimum size of $S$ for which $\rho(S) \geq \nu$. If $T \setminus S = \{j_1, \cdots j_r\}$ , we have (see Eq. (2) in Ref. [22]),

$$\bar{\rho}(T) - \bar{\rho}(S) = \rho(T) - \rho(S) = \sum_{t=1}^{r} \rho_{j_t}(S \cup \{j_1, \cdots j_{t-1}\}). \tag{20}$$

Therefore,

$$\bar{\rho}(T) - \bar{\rho}(S) \leq \rho_{j_1}(S) + \kappa\rho_{j_2}(S) + \cdots \kappa^{t-1}\rho_{j_r}(S). \tag{21}$$

Suppose $\bar{\rho}(T) = 1$, for example if $T$ is a vertex cover. Define $\gamma$ to be $\gamma = \max\{\rho_{j_t}(S) : S \subset T, t = 1 \cdots r\}$. We can get a lower bound on the rank of $S$ using Eq. (21) and the inequality $0 \leq \rho_j(S) \leq \gamma$. First assume $\gamma$ is known. We know that if $S \neq \emptyset$, then $\gamma < 1$. Then,

$$\bar{\rho}(S) \geq 1-\gamma\sum_{t=1}^{r}\kappa^{t-1}. \tag{22}$$

Let us now suppose that:

$$(1-\gamma\sum_{t=1}^{r}\kappa^{t-1}) \geq \nu, \tag{23}$$

and $|S| \geq m$. If an approximation of quality $\nu$ is required, and $r(\nu)$ is the largest value of $r$ such that inequality (Eq. (23)) holds, then $r \leq r(\nu)$. Now $K = C-r$ is the cardinality of $S$ so that $C-r(\nu) \leq C-r$. Thus, the smallest possible value of $|S|$ is,

$$m(\nu) = C - r(\nu). \tag{24}$$

In particular any $m$ must satisfy $m \geq m(\nu)$. Conversely, given $m$, the quality of the approximation depends on $\gamma$ and $r = C-m$. More precisely, the largest value of $\nu$ and thus the largest guaranteed quality of an approximation obtained by our method, has an upper bound given by the left hand side of inequality (Eq. (23)).

## 5. A Formula for the Increment in *F*

This section contains a derivation of the increment $F(S)-F(S\cup\{u\})$ in terms of the first hitting time to $S$ and to $u$. It provides some intuitive insight into the non-increasing and supermodular properties of $F$. Our approach is inspired by the analysis of consensus models [16]. It is very convenient to use the variables $\tau_A$, the first time a random walker visits the set $A \subset V$ and $\tau_A^i$, the time a random walker starting at $i \in V$, visits $A$ for the first time. The main result is a representation of the decrease in $F$ due to adding an element $u$ to the set $S$.

**Theorem 1** *If $S \subset V$, and $u \notin S$ is adjoined to $S$, the decrease in $F$ is*

$$F(S)-F(S\cup\{u\}) = \mathbb{E}_u[\tau_S] + \sum_{i \notin S\cup\{u\}} \mathbb{P}_i[\tau_u < \tau_S]\cdot \mathbb{E}_u[\tau_S]. \tag{25}$$

This result is a direct consequence of Theorem 2 stated at the end of the section. We now state a series of short lemmas that lead to its proof. Given $\Omega$, the sample space of the Markov chain defining the random walk, we can describe the statistics of the first hitting times in terms of an element $\omega \in \Omega$. Each such $\omega$ is associated with a sample path of the Markov chain.

**Definition 2** *Let $S \subseteq V$ and suppose $i, j$ and $u$ are nodes in $V \setminus S$. The subsets $E_{ij}^l(S), F_{ij}^l(S,u) \subset \Omega$ are defined as*

- $E_{ij}^l(S) = \{\omega \mid X_0(\omega) = i, X_r(\omega) \notin S, 0 \leq r \leq l, X_l(\omega) = j \in V\setminus S\}$.
- $F_{ij}^l(S,u) = E_{ij}^l(S) \cap \bigcup_{m=0}^{l}\{\omega \mid \tau_u^i(\omega) = m\}$.

**REMARK:** The definition of $F_{ij}^l(S,u)$ given here uses first hitting times rather than regular times as in Sec. 3. Nevertheless, the sets are identical. It is not hard to show that these sets have the following properties:

**Lemma 3** *Let* $S \subset V$ *and suppose* $i, j$ *and* $u$ *are nodes that are not in* $S$. *Then*

- $E_{ij}^l(S) = E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S,u)$
- $E_{ij}^l(S \cup \{u\})$ and $F_{ij}^l(S,u)$ are disjoint

**Proof:** By definition $\omega \in E_{ij}^l(S)$ if and only if $X_r(\omega) \notin S$ during the time interval $0 \le r \le l$. If the random walk visits $u$ during $[0,l]$ then $\tau_u^i \le l$. Thus $\omega \in F_{ij}^l(S,u)$. If the walk does not visit $u$ then $\omega \in E_{ij}^l(S \cup \{u\})$. Thus $E_{ij}^l(S) \subseteq E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S,u)$. On the other hand the definition implies $F_{ij}^l(S,u) \subseteq E_{ij}^l(S)$ and $E_{ij}^l(S \cup \{u\}) \subseteq E_{ij}^l(S)$. Therefore $E_{ij}^l(S) \supseteq E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S,u)$, establishing the first property. Any $\omega \in E_{ij}^l(S)$ corresponds to a path that either visits $u$ during $[0,l]$ or it does not. Thus the sets $F_{ij}^l(S,u)$ and $E_{ij}^l(S \cup \{u\})$ are disjoint. □

For a subset $W \subset \Omega$, denote its indicator function by $\mathbf{1}_W$. We have the following consequence of Lemma 3.

**Lemma 4** *The indicator function of the sets in Lemma 3 satisfy*

$$\mathbf{1}_{E_{ij}^l(S)} = \mathbf{1}_{E_{ij}^l(S \cup \{u\})} + \mathbf{1}_{F_{ij}^l(S,u)}. \tag{26}$$

**Proof:** Since the sets corresponding to the right hand side of Eq. (26) are disjoint (Lemma 3), we have $\mathbf{1}_{E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S,u)} = \mathbf{1}_{E_{ij}^l(S \cup \{u\})} + \mathbf{1}_{F_{ij}^l(S,u)}$. Thus the value of the right hand side of Eq. (26) is either $0$ or $1$. The left hand side of Eq. (26) is the characteristic function of the set $E_{ij}^l$ while the right hand side is the characteristic function of the set union $E_{ij}^l(S \cup \{u\}) \cup F_{ij}^l(S,u)$. Thus by Lemma 3, the left hand side of Eq. (26) is $1$ if and only if the right hand side is $1$. □

By summing the left hand side and right hand side of Eq. (26) over all $j \notin S$ and then summing over all $l \ge 0$, one obtains for each fixed $i, u \notin S$, an expression for the difference $\mathbb{E}_i[\tau_S] - \mathbb{E}_i[\tau_{S \cup \{u\}}]$. Here $i \ne u$. The representation of $F(S) - F(S \cup \{u\})$ then follows by summing over $i \notin S$. That is the substance of the exposition below.

**Lemma 5** *Let* $S \subset V$ *and* $i$, $j$, *and* $u$ *are nodes in* $V \setminus S$, *with* $i \ne j$, *and* $u \ne i, j$. *Then*

$$\mathbb{E}_i[\tau_S] = \sum_{l=0}^{\infty} [\sum_{j \in V \setminus S} \mathbf{1}_{E_{ij}^l(S)}] \tag{27}$$

**Proof:** Beginning with the fact that $\mathbb{E}[\sum_{j \in V \setminus S} \mathbf{1}_{E_{ij}^l(S)}] = \sum_{j \in V \setminus S} \mathbb{E}[\mathbf{1}_{E_{ij}^l(S)}]$, let us next observe that $\mathbb{E}[\mathbf{1}_{E_{ij}^l(S)}] = \mathbb{P}_i[X_r \notin S, 0 \le r \le l, X_l = j]$. Thus,

$$\sum_{j \in V \setminus S} \mathbb{E}[\mathbf{1}_{E_{ij}^l(S)}] = \mathbb{P}_i[\tau_S > l] = \mathbb{P}_i[\tau_S \ge l+1]. \tag{28}$$

After summing both sides of Eq. (28) over $l \ge 0$, we find that the left hand side is $\mathbb{E}_i[\tau_S]$. □.

**NOTE:** Since we have a finite irreducible Markov chain, $\mathbb{E}_i[\tau_S]$ is finite so the infinite series in $l$ converges.

**Lemma 6** $\sum_{l=0}^{\infty}\mathbb{E}[\sum_{j\in V\setminus S}\mathbf{1}_{F_{ij}^l(S,u)}] = \sum_{l=0}^{\infty}\sum_{k=0}^{l}\mathbb{P}_i[\tau_u = k, \tau_S > l]$

**Proof:** As in lemma 5, $\mathbb{E}[\sum_{j\in V\setminus S}\mathbf{1}_{F_{ij}^l(S,u)}] = \sum_{j\in V\setminus S}\mathbb{E}[\mathbf{1}_{F_{ij}^l(S,u)}]$. Now,

$$\mathbb{E}[\mathbf{1}_{F_{ij}^l(S,u)}] = \mathbb{P}_i[F_{ij}^l(S,u)]$$

Thus,

$$\sum_{j\in V\setminus S}\mathbb{E}[\mathbf{1}_{F_{ij}^l(S,u)}] = \sum_{k=0}^{l}\mathbb{P}_i[\tau_u = k, \tau_S > l]. \tag{29}$$

After summing both sides of Eq. (29) over $l \geq 0$, one obtains

$$\sum_{l=0}^{\infty}\mathbb{E}[\sum_{j\in V\setminus S}\mathbf{1}_{F_{ij}^l(S,u)})] = \sum_{l=0}^{\infty}\sum_{k=0}^{l}\mathbb{P}_i[\tau_u = k, \tau_S > l]. \tag{30}$$

$\square$.

Let us next make the observation:

**Lemma 7** *Let* $\mathcal{X} = \{\omega \mid \tau_S^i(\omega) > l, \tau_u^i(\omega) = k\}$ *and* $\mathcal{Y} = \{\omega \mid \tau_S^i(\omega) > l, \tau_u^i(\omega) = k, k < \tau_S^i(\omega)\}$. *If* $l \geq k$, *then* $\mathcal{X} = \mathcal{Y}$, *with probability one.*

**Proof:** Clearly $\mathcal{Y} \subseteq \mathcal{X}$. Conversely, if $l \geq k$ and $\omega \in \mathcal{X}$ then $\mathcal{X} \subseteq \mathcal{Y}$. $\square$

The conclusion of the following proposition and Lemma 5 applied to the sets $S$ and $S \cup \{u\}$ will allow us to write down a representation of the difference $\mathbb{E}_i[\tau_S] - \mathbb{E}_i[\tau_{S\cup\{u\}}]$

**Proposition 3**

$$\sum_{l=0}^{\infty}\mathbb{E}[\sum_{j\in V\setminus S}\mathbf{1}_{F_{ij}^l(S,u)}] = \mathbb{P}_i[\tau_u < \tau_S]\mathbb{E}_u[\tau_S] \tag{31}$$

**Proof:** Beginning with the right hand side of Lemma 6, we can interchange the order of summation to rewrite it as $\sum_{k=0}^{\infty}\sum_{l=k}^{\infty}\mathbb{P}_i[\tau_u = k, \tau_S > l]$. We then have,

$$\sum_{k=0}^{\infty}\sum_{l=k}^{\infty}\mathbb{P}_i[\tau_u = k, \tau_S > l] = \sum_{k=0}^{\infty}\sum_{l=k}^{\infty}\mathbb{P}_i[\tau_u = k, \tau_S > l, k < \tau_S^i] = \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\mathbb{P}_i[\tau_u = k, \tau_S \geq l, k < \tau_S^i].$$

Here we used Lemma 7. Next we will apply the Strong Markov property [23] to the last expression,

$$\sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\mathbb{P}_i[\tau_u = k, \tau_S \geq l, k < \tau_S^i] = \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\mathbb{P}_i[\tau_S \geq l \mid \tau_u = k, k < \tau_S^i]\cdot\mathbb{P}[\tau_u^i = k, k < \tau_S^i]$$

which is equal to

$$= \sum_{k=0}^{\infty}\sum_{l=k+1}^{\infty}\mathbb{P}_u[\tau_S \geq l-k]\cdot\mathbb{P}[\tau_u^i = k, k < \tau_S^i] = \sum_{k=0}^{\infty}\sum_{l=1}^{\infty}\mathbb{P}_u[\tau_S > l]\cdot\mathbb{P}[\tau_u^i = k, k < \tau_S^i]. \tag{32}$$

On rewriting the right hand side of Eq. (32), we obtain

$$\sum_{k=0}^{\infty}\sum_{l=1}^{\infty}\mathbb{P}_u[\tau_S > l]\cdot\mathbb{P}[\tau_u^i = k, k < \tau_S^i] = \left(\sum_{l=1}^{\infty}\mathbb{P}_u[\tau_S \geq l]\right)\cdot\sum_{k=0}^{\infty}\mathbb{P}[\tau_u^i = k, k < \tau_S^i]. \tag{33}$$

Thus the right hand side of Eq. (33) becomes

$$\mathbb{E}_u[\tau_S]\cdot\mathbb{P}_i[\tau_u < \tau_S]. \tag{34}$$

□

This finally brings us to the formula for the difference in expectations.

**Theorem 2** *Let $S \subset V$ and further let $i, u$ be nodes not in $S$. The relation between the expected hit times $\tau_{S\cup\{u\}}^i$ and $\tau_S^i$ is:*

$$\mathbb{E}[\tau_S] = \mathbb{E}_i[\tau_{S\cup\{u\}}] + \mathbb{P}_i[\tau_u < \tau_S]\cdot\mathbb{E}_u[\tau_S]. \tag{35}$$

**Proof:** The steps in the proof of Lemma 5 can be applied to $S\cup\{u\}$ to prove that $\mathbb{E}_i[\tau_{S\cup\{u\}}] = \sum_{l=0}^{\infty}\mathbb{E}[\sum_{j\in V\setminus S}\mathbf{1}_{E_{ij}^l(S)}]$. Using Lemma 4, Eq. (26), and the convergence of the resulting infinite series we have,

$$\sum_{l=0}^{\infty}\mathbb{E}_i\left[\sum_{j\in V\setminus S}\mathbf{1}_{E_{ij}^l(S)}\right] = \sum_{l=0}^{\infty}\mathbb{E}\left[\sum_{j\in V\setminus S}\mathbf{1}_{E_{ij}^l(S)}\right] + \sum_{l=0}^{\infty}\mathbb{E}\left[\sum_{j\in V\setminus S}\mathbf{1}_{F_{ij}^l(S,u)}\right]. \tag{36}$$

The conclusion of the theorem now follows from Proposition 3 and Lemma 5. □.

## 6. 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 [16] and [13] 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 consider the problem in Eq. (1) over all subsets of cardinality up to $K$, we restricted the search for an optimizing set to classes of optimal and near optimal sets that are closed under the addition and deletion of elements. A ranking function $\bar{\rho}$ is used to rank subsets of $V$ relative to a vertex cover of the graph whose cardinality $C$, satisfies $C > K$. For a constant

$\nu > 0$ the class $L_{\nu,C}$ of near optimal and optimal sets have rank greater than or equal to $\nu$. Let $m$ be the minimum cardinality of sets in this class. Our method consists of first selecting a collection of sets $\mathbf{S}$ of cardinality $m$. The offered approximation is then the best set that results from the greedy extension of each set in $\mathbf{S}$ to a set of cardinality $K$. We discuss two choices for $\mathbf{S}$. The first is based on the observation that a vertex cover is an optimal set for its cardinality and often contains optimal subsets. Figure 2 shows a graph with $N = 9$ vertices and Fig. 1 displays the results of applying the method to a vertex cover. We note that the method also finds optimal solutions which are not subsets of the cover. A second approach does not use a vertex cover. Indeed, there are vertex covers that contain no optimal sets as proper subsets. In this situation one can construct a class of sets that are closed under addition and deletion of certain elements because they are the feasible sets of a greedoid (see Sec. 3). In this situation $\mathbf{S}$ is the set of feasible sets of cardinality $m$, and they are in $L_{\nu,C}$, moreover, all supersets and thus all greedy extensions of $\mathbf{S}$ are also in $L_{\nu,C}$. Our offered approximation is guaranteed in both approaches to be optimal or near optimal. Somewhat more can be said. If the rank of the solution obtained by the classic greedy method is less than $\nu$ or if the solution at stage $m$ is in $\mathbf{S}$, then $S^*$, the solution obtained by our method satisfies, $\rho(S^*) \geq (1+\chi)(1-\frac{1}{e})\rho(\mathcal{O}_K^*)$ where $\mathcal{O}_K^*$ is an optimal solution of problem as posed in Eq. (1) and $\chi > 0$ is a constant that can be computed once the ranks of $S^*$ and the greedy solution are known. The work of Ref. [14] shows that the greedy solution satisfies this inequality for the value $\chi = 0$. In this sense our solution is guaranteed to be at least as good as the greedy solution and will often be better as illustrated by the results of the method for the graph in Fig. 2 where exact solutions are obtained while the greedy solutions are approximate. Our method has computational complexity $O(N^{m+3})$ and in Sec. 4.2 we discuss the tradeoff between computational effort as measured by $m$ and solution quality $\nu$. The identification of classes of graphs for which $m$ is small and where information about $\chi$ can be obtained in advance is an important area for future research.

## 7. References


[1] Olfati-Saber R, Fax JA, Murray RM (2007) Consensus and Cooperation in Networked Multi-Agent Systems. *Proc IEEE* 95(1):215-233. http://dx.doi.org/10.1109/JPROC.2006.887293

[2] Rao A, Ratnasamy S, Papadimitriou C, Shenker S, & Stoica I (2003) Geographic routing without location information. in *Proceedings of the 9th annual international conference on Mobile computing and networking* (ACM, San Diego, California), pp 96-108.

[3] Kempe D, Kleinberg J, Tardos E (2003) Maximizing the spread of influence through a social network. *Proceedings of the Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, Washington, D.C.), pp 137-146.

[4] Richardson M & Domingos P (2002) Mining knowledge-sharing sites for viral marketing. *Proceedings of the Eighth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (ACM, Edmonton, Alberta, Canada), pp 61-70.

[5] Borgs C, Brautbar M, Chayes J, & Lucier B (2014) Maximizing social influence in nearly optimal time. *Proceedings of the Twenty-Fifth Annual ACM-SIAM Symposium on Discrete Algorithms* (SIAM, Portland, Oregon), pp 946-957.

[6] Giakkoupis G (2014) Tight bounds for rumor spreading with vertex expansion. *Proceedings of the 25th ACM-SIAM Symposium on Discrete Algorithms (SODA)* (SIAM, Portland, Oregon), pp 801-815.

[7] Boyd S, Ghosh A, Prabhakar B, Shah D (2006) Randomized gossip algorithms. *IEEE Trans Inf Theory* 52(6):2508-2530. http://dx.doi.org/10.1109/TIT.2006.874516

[8] Censor-Hillel K, Haeupler B, Kelner J, Maymounkov P (2012) Global computation in a poorly connected world: fast rumor spreading with no dependence on conductance. *Proceedings of the Forty-Fourth Annual ACM Symposium on Theory of Computing* (ACM, New York), pp 961-970.

[9] Jadbabaie A (2004) On geographic routing without location information. *43rd IEEE Conference on Decision and Control* (IEEE Control Systems Society), pp 4764-4769.

[10] Krause A, Leskovec J, Guestrin C, VanBriesen J, Faloutsos C (2008) Efficient sensor placement optimization for securing large water distribution networks. *J Water Res Pl-Asce* 134(6):516-526. http://dx.doi.org/10.1061/(ASCE)0733-9496(2008)134:6(516)

[11] Lambiotte R, Sinatra R, Delvenne JC, Evans TS, Barahona M, Latora V (2011) Flow graphs: interweaving dynamics and structure. *Phys Rev E* 84(1). http://dx.doi.org/10.1103/Physreve.84.017102

[12] Kemeny JG & Snell JL (1976) *Finite Markov chains* (Springer-Verlag, New York).

[13] Borkar VS, Nair J, Sanketh N (2010) Manufacturing consent. *48th Annual Allerton Conference* (Allerton House, UIUC, Illinois), pp 1550-1555.

[14] Nemhauser GL, Wolsey LA, & Fisher ML (1978) An analysis of approximations for maximizing submodular set functions—I. *Mathematical Programming* 14(1):265-294. http://dx.doi.org/10.1007/bf01588971

[15] Sviridenko M, Vondrák J, Ward J (2015) Optimal approximation for submodular and supermodular optimization with bounded curvature. *Proceedings of the Twenty-Sixth Annual ACM-SIAM Symposium on Discrete Algorithms (SODA15)* (SIAM, San Diego, California), pp 1134-1148.

[16] Clark A, Bushnell L, Poovendran R (2012) 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, Paderborn, Germany), pp 111-115.

[17] Korte B, Schrader R, Lovász L (1991) Greedoids. *Algorithms and Combinatorics*, eds Cook WJ, Graham RL, Korte B, Lovász L, Wigderson A, Ziegler GM (Springer-Verlag, Berlin, Germany), Vol 4.

[18] Hunt FY (2014) *The structure of optimal and near optimal target sets in consensus models*, (U.S. Department of Commerce, Washington, D.C.) NIST Special Publication 500-303. http://dx.doi.org/10.6028/NIST.SP.500-303

[19] Cormen TH (2009) *Introduction to Algorithms* (MIT Press, Cambridge, MA), 3rd Ed.

[20] Björner A & Ziegler GM (1992) Introduction to Greedoids. *Matroid Applications,* Encyclopedia of Mathematics, ed White N (Cambridge University Press, London, UK), Vol 40.

[21] Il'ev VP (2001) An approximation guarantee of the greedy descent algorithm for minimizing a supermodular set function. *Discrete Appl Math* 114(1–3):131-146. http://dx.doi.org/10.1016/S0166-218X(00)00366-8

[22] Wang Z, Moran B, Wang X, Pan Q (2016) Approximation for maximizing monotone non-decreasing set functions with a greedy method. *J Comb Optim* 31(1):29-43. http://dx.doi.org/10.1007/s10878-014-9707-3

[23] Grimmett G & Stirzaker D (2001) *Probability and Random Processes* (Oxford University Press, New York), 3rd Ed.

***About the author:*** *Fern Hunt is a mathematician in the Applied and Computational Mathematics Division in the Information Technology Laboratory. Dr. Hunt uses tools from probability theory to investigate the influence of network topology on the propagation of information and/or infection in complex networks. The National Institute of Standards and Technology is an agency of the U.S. Department of Commerce.*