early stages of many important threads, but are involved in very few unimportant threads. If we are able to identify such "bellwether" users, then tracking these users can quickly point people who make policies to emerging important threads that are in their early stages of evolution, without inundating them with irrelevant information.

A similar problem has been considered by Leskovec et al. [65] for the placement of contamination sensors in a water distribution network. They show that this approach can be extended to the unseemingly similar problem of selecting a set of blogs to monitor so as to catch the maximum number of important news stories. More recently El-Arini et al. [34] considered a similar problem of providing personalized monitoring of the blogosphere, tailored to individual users. In both cases these problems can be formalized as instances of *submodular maximization problems* that have a long history starting with Nemhauser et al. in 1978 [78]. For set U the function $f : 2^U \rightarrow \mathbb{R}^+$ is said to be submodular if it exhibits the property of "diminishing returns": $f(A) + f(B) \geq f(A \cup B) + f(A \cap B)$ for all $A, B \subseteq S$. Nemhauser et al. [78] consider the problem of finding a subset $S \subseteq U$ of given cardinality which maximizes $f(S)$ and show that a simple greedy algorithm that continually adds to S the element $u \in U$ which maximizes $f(S \cup \{u\}) - f(S)$ provides a near-optimal solution. Khuller et al. [55] extend this result to a problem *budgeted maximum coverage* where each element $u \in U$ has associated cost c_u and the objective is to find subset $S \subset U$ such that $f(U)$ is maximized and $\sum_{u \in S} c_u \leq b$ for some budget b. More advanced constraints have also been shown to have near-optimal approximations [21, 20].

In Chapter 5 we show that the problem of determining "bellwether" users on