STRATEGIC INFORMATION DISSEMINATION AND LINK FORMATION IN SOCIAL NETWORKS

Yu Zhang, Mihaela van der Schaar

Electrical Engineering Department, University of California, Los Angeles

ABSTRACT

In this paper, we propose a novel game-theoretic framework for analyzing and understanding how strategic networks are formed endogenously, driven by the self-interested decisions of individual agents aiming to maximize their own utilities by trading-off the costs and benefits of forming links with other agents. We explicitly model and analyze the scenario in which agents benefit from disseminating their own information to other agents. We rigorously prove that the equilibria of strategic networks frequently exhibit a core-periphery structure, where there are only few agents at the center (core) of the network while the majority of agents are at the periphery of the network and communicate with other agents via links maintained by the "core" agents, who play the role of "connectors" in the network. Also, we are able to determine under what conditions the strategic networks operating in equilibrium are minimally connected (i.e. there is a unique path between any two agents) and have short network diameters. These properties are commonly observed on the Internet and important because they ensure the efficiency and robustness of the resulting equilibrium networks. However, none of these has been rigorously proven in a formal framework before.

Index Terms— Strategic networks, information dissemination, link formation

1. INTRODUCTION

In traditional networks (e.g. communication networks), agents are obedient and the topology of the network is dictated by a system designer [3][4]. Different from this, a common feature of the emerging social networks, e.g. social networks like Twitter [1] and expert networks like Amazon Mechanical Turk [2], is that agents (i.e. people or smart machines) are usually self-interested and can proactively create and dissolve links to other agents in order to maximize their own utilities from the information exchange. In this paper, we refer to such networks, formed by the agents' self-interested (strategic) decisions, as *strategic networks*. This strategic behavior of the agents shapes the emerging network topologies as well as their stability and efficiency. A novel framework is thus necessary to enable the *rigorous study* of the efficiency and the stability of such strategic networks.

The theoretical study on link formation in social and economic networks has been conducted by microeconomics researchers as well as engineers[7][8], who analyze the emerging network topologies under the agents' self-interest. These works focus on the scenario in which agents benefit solely from consuming information produced by other agents (e.g. downloading files in P2P networks). In this scenario, the agents' benefits only depend on the total amount of information that they consume [8]. The scenario where agents' benefits come from disseminating their own information to other agents (e.g. advertising in social networks [1]), was not considered in the existing literature. In this work, we specifically focus on the scenario where agents' benefits come from to information dissemination, which we refer to as the Information Dissemination Game (IDG). In the IDG, since the benefit of each agent comes from the information dissemination to others, the information possessed by other agents has no influence on its incentive. In contrast, the number and variety of agents it connects with form the most important factor that shapes its incentive, which exhibit a sharp contrast to the existing literature.

We consider a strategic network whose topology is constructed by agents strategically forming links to disseminate their information. Each agent is able to disseminate its information to any other agent connected via a path. The benefit of an individual agent comes from the reception of its disseminated information by other agents. During the link formation process, each agent wants to maximize its own benefit from information dissemination over the formed links while minimizing its link formation cost. Given the formalism of the IDG, we study what non-cooperative equilibria (i.e. equilibria resulting based on the agents' self-interested link formation actions) emerge and characterize how efficient these equilibria are from a social welfare perspective compared to the socially optimal networks.

Our analysis is able to characterize and prove several important properties of the strategic network at equilibria, which have been empirically measured in the network science literature [5][6], but have not been rigorously proven in a formal framework. First, we show that given the agents' interests in information dissemination, strategic networks operating in equilibrium usually exhibit core-periphery topologies where there are only a few agents at the center (core) of the network while the majority of agents are at the periphery of the network and communicate with other agents via links maintained by the core agents. Also, the proposed framework rigorously proves that the strategic networks operating in equilibrium are often minimally connected (i.e. there is a unique path between any two agents), and we prove that they have short network diameters which are independent of the population size of the network. We also briefly discuss the dynamic link formation process where agents can learn the strategies played by others and asynchronously adapt their strategies by playing the best response dynamics. In summary, the goal of our framework and analysis is to guide network and application designers in understanding the impact of strategic interactions among agents and enable them to design better incentive protocols that enhance the stability and efficiency of the network.

The remainder of this paper is organized as follows. In Section 2, we describe our model of the IDG. In Section 3, we characterize the properties of the emerging non-cooperative equilibria in IDG. Section 4 studies the dynamic link formation process. Section 5 concludes the paper.

2. SYSTEM MODEL

Let $N = \{1, 2, \ldots, n\}$ be the set of agents in the network with $n \geq 3$ and let i and j denote typical members in this set. Each agent i possesses some information (e.g. news, advertisement, data) in the amount $x_i \in \mathbb{R}^+$ which it finds in its own benefit to disseminate to other agents. We consider a simultaneous move game where agents strategically determine whether to create links with other agents in order to disseminate their information. We consider unilateral link formation, where links are created by the unilateral actions of agents and link costs are one-sided. Thus, the mutual consent of two agents in order to create a link between them is not required. The link formation strategy adopted by an agent i is denoted by a tuple $\boldsymbol{g}_i = (g_{ij})_{j \neq i} \in \{0,1\}^{n-1}$. Each element g_{ij} in this tuple is binary and indicates whether agent i formed a link with agent $j \in N / \{i\}$. We set $g_{ij} = 1$ if agent i forms a link to agent j, and $g_{ij} = 0$ otherwise. The creation of a link incurs a cost to the creator and hence, its decision to form a link involves trading-off the benefit received from disseminating information using this link and the incurred cost. Given the strategies of agents, a strategy profile in the information dissemination game is defined as $\boldsymbol{g} \triangleq (\boldsymbol{g}_i)_{i=1}^n$. It should be noted that since g_{ij} is binary and n is finite, the space which $\boldsymbol{g}$ takes value from is also finite, which is denoted as $\boldsymbol{G}$.

The information flow across a link is assumed to be undirected. That is, given a link between any two agents, the information can be transmitted in both directions (i.e. from the creator to the recipient and vice versa) across this link. We define $\overline{\boldsymbol{g}} \triangleq (\overline{\boldsymbol{g}}_i)_{i=1}^n$ to indicate the agents' connectivity, with $\overline{\boldsymbol{g}}_i = (\overline{g}_{ij})_{j \neq i}$ and $\overline{g}_{ij} = \max\{g_{ij}, g_{ji}\}$ for each $j \in N / \{i\}$. We can thus define the *topology* of the network as the closure of $\overline{\boldsymbol{g}}$, i.e. $cl(\overline{\boldsymbol{g}}) = \{(i,j) \in N \times N \mid i \neq j,\ \overline{g}_{ij} = 1\}$.

It can be easily proved that there will be at most one link between any two agents i, j upon their self-interests, since the information flow is undirected and hence it is wasteful for an agent to redundantly form an already established link. Therefore, we use (i,j) and (j,i) refer to the same link formed by agent i if $g_{ij} = 1$, and vice versa. Hence, $cl(\overline{\boldsymbol{g}})$ also represents an undirected network. In the rest of this paper, we will use the terms "topology" and "network" interchangeably. A *path* in a network is defined as follows.

Definition 1 (Path). Given an network $cl(\overline{\boldsymbol{g}})$, a path between agents i and j is a tuple $path_{ij} = ((i,j_1),(j_1,j_2),\ldots,(j_m,j))$ for some $m \geq 0$, where $j_1,\ldots,j_m$ are agents distinct from i and j such that $path_{ij} \subseteq cl(\overline{\boldsymbol{g}})$.

Two agents i and j are called *connected* in a network $cl(\overline{\boldsymbol{g}})$, if and only if there is at least one path between them in $cl(\overline{\boldsymbol{g}})$. We assume that an agent i can disseminate its information to any agent j with whom it has a path. The benefit that agent i receives by disseminating its information to agent j is proportional to the amount of information x_i. Given this, the utility of an agent i in the IDG can be expressed as:

$$u_i(\boldsymbol{g}) = x_i \sum_{j \in N_i(\overline{\boldsymbol{g}})} \alpha_{ij} - \sum_{j \in N_i(\boldsymbol{g})} k_{ij}\ . \qquad (1)$$

Here $\alpha_{ij} \in \mathbb{R}^+$ is the benefit that agent i receives by disseminating unit information to agent j and $k_{ij} \in \mathbb{R}^+$ is the link formation cost for agent i to form a link with agent j. $N_i(\overline{\boldsymbol{g}})$ is defined as the set of agents to whom agent i is connected, and $N_i(\boldsymbol{g})$ is defined as the set of agents to whom agent i forms links. Without loss of generality, we normalize $\alpha_{ij} \in [0,1]$. Below we briefly explain the utility function in (1).

Remark 1: We assume that when there are multiple paths between agent i and agent j and multiple copies of agent i's information are received by agent i, agent i receives a fixed benefit of $x_i \alpha_{ij}$ regardless of the number of copies that agent j receives. We also assume that each agent can only benefit from disseminating its own information, but forwarding the information that is received from other agents does not bring it any benefit.

In the IDG, each agent maximizes its own utility given the strategies of others. A Nash equilibrium (NE) is defined as a strategy profile $\boldsymbol{g}^*$ such that the strategy of each agent i is a best response to the strategies of others:

$$u_i(\boldsymbol{g}_i^*, \boldsymbol{g}_{-i}^*) \geq u_i(\boldsymbol{g}_i', \boldsymbol{g}_{-i}^*), \forall \boldsymbol{g}_i' \neq \boldsymbol{g}_i^*,\ \forall i \in N\ . \ (2)$$

Here we use $\boldsymbol{g}_{-i}$ to represent the strategies of all agents other than agent i. In addition, a strict NE is a Nash equilibrium such that the strategy of each agent i is a strict best response to the strategies of others (with the inequality in (2) being strict for each $i \in N$). As we will show in Section 4, a network will always converge to a strict NE in the dynamic link formation process, given its Therefore, a strict NE characterizes a steady state in the dynamic link formation process. However, it should also be noted that since strict NE are a subset of NE, all our subsequent results on NE also apply to strict NE.

The social welfare of the IDG is defined to be the sum of agents' individual utilities. For a strategy profile $\boldsymbol{g}$, the social welfare is given by $U(\boldsymbol{g}) \triangleq \sum_{i \in N} u_i(\boldsymbol{g})$. A strategy profile $\boldsymbol{g}^{\#}$ is called socially optimal if it achieves the social optimum, denoted by $U^{\#}$, i.e. $U^{\#} \triangleq U(\boldsymbol{g}^{\#}) \geq U(\boldsymbol{g}'),\ \forall \boldsymbol{g}' \neq \boldsymbol{g}^{\#}$.

3. EQUILIBRIUM AND EFFICIENCY

In this section, we first analyze the equilibrium link formation strategies of self-interested agents. Next, we explicitly compare the social welfare of the IDG at equilibrium with the social optimum. The results provide important insights on the efficiency loss occurred due to the self-interested behavior of agents in the IDG as compared to the case when the agents obediently follow the link formation actions dictated by some central designer.

3.1 Equilibrium analysis

We start by defining several useful concepts for our analysis.

Definition 2 (Component). A component C is a set of agents such that for any $i, j \in C$, they are connected, while for any $i \in C$ and $j' \notin C$, they are not connected.

Remark 2: It should be noted that an agent who is not connected with any other agents in the network (i.e. an isolated agent) also forms a component, which is called as a *singleton* component. Accordingly, a component that is not singleton is called a *non-singleton* component. Also, a component is called to be *minimal* if and only if there is only one path in $cl(\overline{\boldsymbol{g}})$ between any two agents $i, j \in C$. Given a component, the *distance* $d_{\boldsymbol{g}}(i,j)$ between any two agents i and j in it is defined as the number of links in a shortest path between them. The diameter of a component is thus defined as the largest distance between any two agents in it, and the diameter of the network is defined to the largest diameter of its components.

For better illustration, we also define several particular networks. A network $cl(\overline{\boldsymbol{g}})$ is called *connected* if there is a unique component $C = N$ in $cl(\overline{\boldsymbol{g}})$. A network $cl(\overline{\boldsymbol{g}})$ is called *empty*, denoted as $\boldsymbol{g}^e$, if $cl(\overline{\boldsymbol{g}}) = \phi$. A network $cl(\overline{\boldsymbol{g}})$ is called a star network if there is an agent $i \in N$ such that $\overline{g}_{ij} = 1, \forall j \in N / \{i\}$ and $\overline{g}_{jj'} = 0, \forall j, j' \in N / \{i\}$.

It is easy to prove that Nash equilibria always exist in the IDG with the formal proof omitted here and relegated to the online appendix [9]. In the following proposition, it is further shown that under an NE of the IDG, each component is minimal. It should be noted that due to the space limit, proofs are omitted and can be found in the online appendix [9].

Proposition 1. Under an NE, each component is minimal. ■

Proposition 1 shows that in an equilibium, agents will not make redundant investments on the link formation and hence, each connected sub-network (component) in the network is minimal with no cycles in it. As we will show in Section 3.2, the social optimum in the IDG is always achieved by networks consisting of minimal components and hence, the minimal property guarantees that the equilibia can frequently achieve the social optimum.

Proposition 1 characterizes the components in the equilibrium network. However, it does not characterize under what conditions the network is connected, i.e. there is a unique component in the network. The following proposition provides a sufficient condition under which the network is connected.

Proposition 2. The network under each NE is always minimally connected if $\min_{i,j}\{x_i \alpha_{ij} / k_{ij}\} > 1$. ■

Proposition 2 shows that the network will be connected at equilibrium when (i) the link formation cost is not too large; (ii) the benefit from information is sufficiently large. The properties of the network topology at equilibrium (e.g. the shape of the topology) depends on the specific values of $\{x_i\}_{i \in N}$, $\{\alpha_{ij}\}_{i,j \in N}$ and $\{k_{ij}\}_{i,j \in N}$. In the rest of this section, we analyze two exemplary networks with particular structures in order to obtain further insights on the equilibrium topology.

3.1.1 Networks with recipient-dependent costs

In the first example, we consider the network where $k_{ij} = k_j, \forall i \in N / \{j\}$, i.e. the cost of forming a link is exclusively recipient specific. This can capture the practical networks in which the link formation cost only depends on the type of the recipient. For the tractability of the analysis, we assume that $\alpha_{ij} = \alpha, \forall i, j$ and focus solely on the heterogeneity of the link formation cost. In the following proposition, we show that if the link formation cost in the network is not arbitrary but only takes values from a finite set $\{k^1, ..., k^L\}$, i.e. there are L different types of link formation costs and $k_{ij} \in \{k^1, ..., k^L\}, \forall i, j \in N$, then the distance between any two agents within the same component at equilibrium should be no more than $2L + 2$.

Theorem 1. If $k_i \in \{k^1, ..., k^L\}, \forall i \in N$, then under a strict NE $\boldsymbol{g}^*$, the distance between any two agents within the same component is at most $2L + 2$. ■

The link formation cost thus plays an important role in shaping the equilibrium network. If there are only a finite number of different link formation costs in the network, then the size of each component (measured by the longest distances between agents) cannot be arbitrarily large but is upper-bounded by some constant value which is independent to the population size. Based on Proposition 1 and Theorem 1, the "minimally connected" and "short diameter" properties of the equilibria in strategic networks are proven.

As a special case, we prove in the next corollary that when the link formation cost is the same for all agents, each component in a strict NE is a star, regardless of the values $\{x_i\}_{i \in N}$ and $\{\alpha_{ij}\}_{i,j \in N}$.

Corollary 1. If $k_{ij} = k, \forall i, j \in N$, then under a strict NE $\boldsymbol{g}^*$, each component forms a star topology. ■

Therefore, each component at equilibrium preserves the "core-periphery" property with one agent staying at the center of the component and playing the role of the "connector" who maintains links with all other agents for their information dissemination.

3.1.2 Networks with groups

In the second example we discuss a network where agents are divided into groups and agents within the same group have the same type. The benefits from information dissemination within the same group are higher than the benefits received from information dissemination across groups. Also, the cost of forming links within a group is lower than the cost of forming links across groups. One example of strategic networks where such groups exist is devices or processing nodes located in the same area [10].

Formally, we consider all agents are divided into Z different groups $N_1, ..., N_Z$, such that $N = \cup\{N_z\}_{z=1}^{Z}$ and $N_z \cap N_{z'} = \phi$ for any $1 \leq z < z' \leq Z$. For agents i and j within the same group, the benefit for agent i to disseminate one unit of information to agent j is $\alpha_{ij} = \overline{\alpha}$ and the cost for agent i to form a link to agent j is $k_{ij} = \underline{k}$. Similarly, for agents from different groups, the benefit for disseminating one unit of information between them is $\underline{\alpha} < \overline{\alpha}$ and the cost of forming a link between them is $\overline{k} > \underline{k}$. Here we assume that $x_i = x, \forall i \in N$ to make our analysis tractable. The following theorem characterizes the equilibria in IDG with groups and proves that each non-empty network at equilibrium preserves the "core-periphery" property.

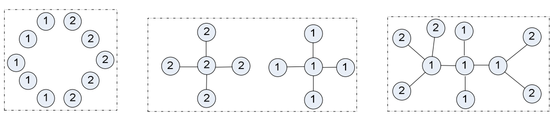
Theorem 2. In the presence of groups, the NE can be characterized as follows:

(i) When $x \in (0, \underline{k} / \overline{\alpha})$, the unique equilibrium is $\boldsymbol{g}^e$;

(ii) When $x \in (\underline{k} / \overline{\alpha}, \overline{k} / \underline{\alpha})$, the unique strict NE consists of Z components, where each component only contains agents from the same group and forms a star topology;

(iii) When $x \in (\bar{k} / \underline{\alpha}, \infty)$, each strict NE contains a group N_z and an agent $i \in N_z$ with $g^*_{ij} = 1, \forall j \in N_z / \{i\}$. For each agent $j' \notin N_z$, there is an agent $j \in N_z$ such that $g^*_{jj'} = 1$. ■

Several examples of the equilibrium topology are illustrated in Figure 1 in a network of $n = 10$ which are divided into 2 groups. The number on the node represents the group that each agent belongs to. Theorem 2 provides several important insights. First, with the exception of empty networks, agents from the same group always belong to the same component under a strict equilibrium. Hence, the heterogeneity on agents' types have significant impact on the resulting equilibrium, where agents of the same type who have high mutual benefits and low connecting cost always get closer with each other than agents of different types. Second, each non-singleton component in the network still preserves the "core-periphery" property at equilibrium. There is one group which is the core with its agents mutually connected, and agents from other groups accessing the network via links maintained by the core group. Third, in each component, there is always a central agent and all paths within this component initiate from this agent. The distance from the central agent to any periphery agent is no more than 2.



(A) $x < \underline{k} / \bar{\alpha}$ (B) $x \in (\underline{k} / \bar{\alpha}, \bar{k} / \underline{\alpha})$ (C) $x > \bar{k} / \underline{\alpha}$

Figure 1 The examplary Nash equilibria in the network with groups

3.2 Equilibrium efficiency

In this section, we analyze the social welfare of the IDG to quantitatively study the efficiency of equilibria. First, we characterize the socially optimal strategy profiles. Similar to the NE, it is easy to prove that each component under a socially optimal strategy profile is also minimal. We then have the following proposition which characterizes the socially optimal network.

Proposition 3. The network under a socially optimal profile is always minimally connected if $\max_{i \in N}\{\min_{j \in N/\{i\}} x_i \alpha_{ij} / k_{ij}\} > 1$. ■

It should be noted that $\max_{i \in N}\{\min_{j \in N/\{i\}} x_i \alpha_{ij} / k_{ij}\} > 1$ always holds when $\min_{i,j}\{x_i \alpha_{ij} / k_{ij}\} > 1$. Hence, we have the following theorem which proves that when the link formation cost is the same for every agent, i.e. $k_{ij} = k, \forall i, j \in N$, an NE achieves the social optimum as long as it forms a connected network.

Theorem 3. (i) An NE $\boldsymbol{g}^*$ is socially optimal if there is a unique component N in $cl(\bar{\boldsymbol{g}}^*)$; (ii) when $\min_{i,j}\{x_i \alpha_{ij} / k_{ij}\} > 1$, each equilibrium $\boldsymbol{g}^*$ achieves the social optimum. ■

Theorem 3 provides important insights in quantifying the efficiency of equilibria. It shows that when the link formation cost is not too large, self-interested link formation actions can lead to the socially optimum without incurring any efficiency loss. This is due to the fact that the network is minimally connected at equilibrium and hence, agents form the minimum number of links that is enough to help them disseminate information throughout the network. Hence, the total link formation cost incurred to agents at equilibrium is minimized.

4. DYNAMIC LINK FORMATION

In the above section, we study the IDG as a static game, where all agents make their link formation actions simultaneously. In this section, we study the dynamic link formation in which agents learn the strategies played by others and asynchronously adapt their strategies by playing the best response dynamics. We assume that the time is divided into discrete periods with t being the time stamp. In each period t, each agent i observes the strategy profile $\boldsymbol{g}^{(t-1)}$ played in the previous period and adapts its strategy to the myopic best response $\boldsymbol{g}^*_i$ such that

$$u_i(\boldsymbol{g}^*_i, \boldsymbol{g}^{(t-1)}_{-i}) \geq u_i(\boldsymbol{g}'_i, \boldsymbol{g}^{(t-1)}_{-i}), \forall \boldsymbol{g}'_i \in \{0,1\}^{n-1} \text{ and } \boldsymbol{g}'_i \neq \boldsymbol{g}^*_i. \quad (3)$$

Remark 3: When there are multiple best responses for an individual agent in one period, we assume that each best response strategy has a positive probability to be chosen by the agent.

We analyze the dynamic link formation process for two exemplary IDG models discussed in Section 3. In the analysis, we assume that $\alpha_{ij} = \alpha, k_{ij} = k, \forall i, j$ and $x_i = x > k, \forall i$.

First, we prove that the dynamic link formation process always converges to a strict equilibrium under both IDG models.

Theorem 4. In the IDG with recipient-dependent cost and the IDG with groups, the best response dynamics always converges to a strict equilibrium. ■

Next, we compare the convergence rate in different IDGs. Figure 2 shows that the convergence rate (measured in the number of periods before convergence) of the best response dynamics grows exponentially against the population size n.

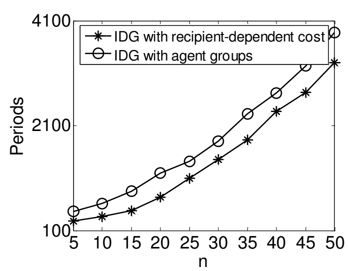


Figure 2 The convergence rate of the dynamic link formation process

5. CONCLUSION

In this work, we investigated the problem of information dissemination and link formation in networks of strategic agents. We rigorously determined how the agents' aspiration to disseminate their own information throughout the network impacts their interactions and the emerging connectivity/topology between them. Our analysis rigorously proved several important properties of the networks emerging in equilibrium from such strategic link formation: "core-periphery", "minimally connected", "short diameter". These properties are important because they characterize that efficiency and robustness of the resulting equilibrium networks. Finally, we proved that agents converge to the characterized equilibrium networks using a simple and asynchronous dynamic link formation process.

6. REFERENCES

[1]. H. Kawk, C. Lee, H. Park and S. Moon, "What is Twitter, a social network or a news media," *Proc. 19th Int'l Conf. on World Wide Web*, 2010.

[2]. A. Kittur, E. Chi and B. Suh, "Crowdsourcing user studies with mechanical turk," *Proc. 26th Annual SIGCHI Conf. on Human Factors in Computing Systems*, 2008.

[3]. T. Melodia, D. Pompili and I. Akyildiz, "On the interdependence f distributed topology control and geographical routing in ad hoc and sensor networks," *IEEE Journal on Sel. Area in Commun.*, vol. 23, no. 3, pp. 520 – 532, 2005.

[4]. M. Papadopouli and H. Schulzrinne, "Effects of power conservation, wireless coverage and cooperation on data disseminataion among mobile devices," *Proc. MobiHoc*, pp. 117 – 127, 2001.

[5]. O. Hein, M. Schwind and W. Konig, "Scale-free networks: The impact of fat tailed degree distribution on diffusion and communication processes," *Wirtschafts Informatik*, vol. 48, no. 4, pp. 267 – 275, 2006.

[6]. D. Watts and S. Strogatz, "Collective dynamics of 'small-world' networks," *Nature*, vol. 393, pp. 440 – 442, 1998.

[7]. J. Kleinberg, S. Suri, E. Tardos and T. Wexler, "Stratetic network formation with structure holes," *ACM SIGecom Exchanges*, vol. 7, no. 3, 2008.

[8]. A. Galeotti and S. Goyal, "The law of the few," *American Economic Review*, vol. 100, no. 4, pp. 1468 – 1492, 2010.

[9]. Y. Zhang and M. van der Schaar, "Online appendix: Strategic Information dissemination and link formation in social networks," *online appendix*, http://medianetlab.ee.ucla.edu/papers/information_dissemination_icassp_appendix.

[10]. Y. Suh, H. Shin and D. Kwon, "An efficient multicast routing protocol in wireless mobile networks," *Wireless Networks*, vol. 7, no. 5, pp. 443 – 453, 2001.