

Article

# Formation of Stable and Efficient Social Storage Cloud

Pramod C. Mane <sup>1</sup>, Nagarajan Krishnamurthy <sup>2</sup> and Kapil Ahuja <sup>1,\*</sup>

<sup>1</sup> Computer Science and Engineering, Indian Institute of Technology Indore, Indore 453552, India; pramod@iiti.ac.in

<sup>2</sup> Operations Management and Quantitative Techniques, Indian Institute of Management Indore, Indore 453556, India; nagarajan@iimidr.ac.in

\* Correspondence: kahuja@iiti.ac.in

Received: 28 June 2019; Accepted: 16 October 2019 ; Published: 1 November 2019



**Abstract:** In this paper, we study the formation of endogenous social storage cloud in a dynamic setting, where rational agents build their data backup connections strategically. We propose a degree-distance-based utility model, which is a combination of benefit and cost functions. The benefit function of an agent captures the expected benefit that the agent obtains by placing its data on others' storage devices, given the prevailing data loss rate in the network. The cost function of an agent captures the cost that the agent incurs to maintain links in the network. With this utility function, we analyze what network is likely to evolve when agents themselves decide with whom they want to form links and with whom they do not. Further, we analyze which networks are pairwise stable and efficient. We show that for the proposed utility function, there always exists a pairwise stable network, which is also efficient. We show that all pairwise stable networks are efficient, and hence, the price of anarchy is the best that is possible. We also study the effect of link addition and deletion between a pair of agents on their, and others', closeness and storage availability.

**Keywords:** network formation; pairwise stability; network externalities; social storage cloud; socially-aware storage-sharing

**MSC:** 91A40; 91A80; 91B32; 91B99

**JEL Classification:** C72; D62; D85; L86

## 1. Introduction

Online data backup services such as BuddyBackup<sup>1</sup> and CrashPlan<sup>2</sup> allow agents to share their under-utilized storage (disk) space with others as well as backup their data on the storage space shared by other agents. In academic discourse, numerous architectural prototypes of data backup systems (for example, Social Storage Cloud [1], Friendstore [2], F2Box [3], FriendBox [4], BlockParty [5], and so on) have been proposed. In order to mitigate issues like data security, trust, or low quality of services, these systems (services) are leveraging social connections. The social connections are either exogenous (that is, encoded in social graphs, for instance, the Facebook social graph<sup>3</sup>) or endogenous (constructed by the agents [6]). Such social connections are at the core of these systems.

<sup>1</sup> <http://www.buddybackup.com> (accessed on 21 June 2019).

<sup>2</sup> <https://support.crashplan.com> (accessed on 21 June 2019).

<sup>3</sup> <https://developers.facebook.com/docs/graph-api> (accessed on 21 June 2019).

A recently published survey [7] in this field mentions various issues related to social connections, such as small friend sets, social closeness quantification, and so on. These issues are discussed in the context of exogenous social connections.

The aspect of endogenous social connections is notably lacking. The agents' self-interested behaviour, guided by the cost-benefit trade-off, in building storage-sharing connections is still poorly understood. Specifically, what is not well understood in this context is: (1) which network structure is likely to emerge when self-interested agents construct their storage sharing connections; (2) whether the emerged storage-sharing connection structure is stable and efficient, or not; and (3) the impact of link formation between two agents on their storage availability as well as that of other agents. To advance our understanding about these aspects, there is a need for formal modeling of endogenous socially-aware storage-sharing networks, as previous studies have focused exclusively on exogenous networks.

This paper studies the aforementioned aspects by focusing on social storage cloud systems (a case of socially-aware resource sharing systems). We model social storage cloud systems as an endogenous social storage cloud by using the tools of *network analysis*<sup>4</sup>, *game theory*<sup>5</sup>, and *network formation*<sup>6</sup>. Specifically, we model social storage cloud systems as a strategic network formation game, where self-interested agents decide with whom they want to form a connection and with whom they do not. For this, we define the utility of agents in a social storage cloud by taking into consideration the parameters data failure rate, value of data, and cost for maintaining social connections.

In [23], the authors consider a degree-based utility model, where agents benefit only from direct neighbors, and the benefit decreases with an increase in the number of neighbors of each neighbour [24]). The utility function we define in this study is degree-distance-based, where agents obtain benefits from direct and indirect neighbors, but the benefit decreases with an increase in the number of direct and indirect neighbors [25]. With this utility function, we study the effect of decisions of addition and deletion of links by pairs of agents on their storage availability in the network. We study externalities in the network, that is, the effect of link formation between a pair of agents on the utility of the other agents. We then analyze the network structure that evolves due to these decisions of link addition and link deletion.

The focus of this paper is to study network stability, efficiency, and the measures of price of anarchy and price of stability. For the analysis of network stability, we make use of the concept of pairwise stability proposed in [26]. In our model, agents experience both positive and negative externalities, determined by storage availability. We provide necessary and sufficient conditions for an agent to experience positive and negative externalities. Further, we show that if data failure rate is less than the ratio of cost of maintaining the link to data value, then the null network is the unique pairwise stable as well as an efficient network. However, if the data failure rate is higher than the ratio of the cost of maintaining the link to data value then a network where every agent has, at most, a single link is the unique pairwise stable and efficient network.

The structure of the paper is as follows. Section 2 discusses the social storage cloud model. Section 3 studies the effect of addition and deletion of a link between a pair of agents on their closeness and storage availability, and that of others. Section 4 discusses the characterization of stable networks, where we study deviation conditions that show when agents have incentives for adding or deleting a link. Further, the section discusses network stability, efficiency, and inefficiency. Section 5 concludes the discussion.

---

<sup>4</sup> We refer the reader to the body of the paper [8–12] for details on centrality measures in networks.

<sup>5</sup> The theoretical game techniques have been quite successfully used in computer science [13–16] and other engineering disciplines [17,18].

<sup>6</sup> The literature on network formation is vast. The surveys [19–22] explore many dimension of this topic.

## 2. Social Storage Cloud Model

In this section, we describe the social storage cloud model through an interaction structure, a storage-sharing framework and cost-benefit analysis of agents.

### 2.1. Interaction Structure

A social storage cloud  $\mathbf{g} = (\mathcal{A}, \mathcal{L})$  is a storage-sharing and data backup network that consists of a nonempty set  $\mathcal{A}$  of  $N$  agents who are involved in storage (disk)-space sharing and data backup activity; and a set,  $\mathcal{L}$ , of links that connect these agents. The set  $\mathcal{L}$  acts as a communication infrastructure for agents to share their storage space with others and search for storage space provided by others. A link,  $\langle ij \rangle \in \mathcal{L}$ , represents a direct communication channel between agents  $i$  and  $j$ , which is bidirectional (and hence,  $\langle ij \rangle = \langle ji \rangle$ ). If  $\langle ij \rangle \in \mathcal{L}$ , we call the agents  $i$  and  $j$  as neighbours in the network  $\mathbf{g}$ . The number of neighbors of agent  $i$  in  $\mathbf{g}$  is denoted by  $\eta_i(\mathbf{g})$ .

Given distinct agents  $a_1, a_2, \dots, a_n \in \mathcal{A}$ , if  $\langle a_1, a_2 \rangle, \langle a_2, a_3 \rangle, \dots, \langle a_{n-1}, a_n \rangle \in \mathcal{L}$ , then there is a path  $\mathcal{P}_{a_1 a_n}(\mathbf{g})$ , from  $a_1$  to  $a_n$ , of length  $n - 1$ . The distance  $d_{ij}(\mathbf{g}) (= d_{ji}(\mathbf{g}))$  between a pair of agents  $i$  and  $j$  is the length of the shortest path connecting them in  $\mathbf{g}$ .

A network  $\mathbf{g}$  is connected if there exists at least one path between any pair of agents, otherwise, it is disconnected. A path of length  $\geq 2$  between a pair of agents is an indirect communication channel between them. The set,  $\mathcal{G}(N)$ , consists of all possible networks on  $N$  agents.

Data stored on local storage space is prone to loss due to multiple reasons such as virus infection, software or hardware failure, data corruption, and so on. Therefore, each agent wants to backup its data on remote storage (disk) space. For any agent, data loss is costly. We capture this by assuming that the value each agent associates with its data is quantifiable and given. Every agent (as a data owner) strives for obtaining storage space provided by other agents (as storage providers) in  $\mathbf{g} \in \mathcal{G}(N)$ . Agent  $i$  wants to backup  $\bar{b}_i$  amount of data and shares  $\bar{s}_i = \sum_{j \in \mathcal{A} \setminus \{i\}} \bar{b}_j$  amount of storage space. This leads to endogenous social storage cloud formation, where each agent builds its communication channel to seek storage space from direct and indirect communication channels. We assume that each agent has global (complete) information about the network structure.

A network  $\mathbf{g}$  evolves when agents perform two actions, namely, link addition ( $\mathbf{g} + \langle ij \rangle$ ) and link deletion ( $\mathbf{g} - \langle ij \rangle$ ). *Mutual consent of a pair of agents is required for addition of a link between them, but any link can be unilaterally deleted.*

Table 1 summarizes all notations used in this paper.

**Table 1.** Notation summary.

$\mathbf{g}$	social storage cloud.
$\mathcal{A}$	set of agents (or vertices).
$N$	the number of elements in the set $\mathcal{A}$ , which is the number of agents in $\mathbf{g}$ .
$\mathcal{L}$	set of links (or edges).
$\langle ij \rangle$	link between agents $i$ and $j$ .
$\zeta$	cost incurred by each agent to maintain a link.
$\lambda$	probability that an agent loses its data.
$\beta$	worth (or value) that each agent has for its data.
$\Phi_i(\mathbf{g})$	closeness of agent $i$ in $\mathbf{g}$ .
$\alpha_{ij}(\mathbf{g})$	probability that agent $i$ obtains storage space from agent $j$ in $\mathbf{g}$ .
$\gamma_i(\mathbf{g})$	probability that agent $i$ obtains storage space from at least one agent in $\mathbf{g}$ .
$\eta_i(\mathbf{g})$	neighborhood size of agent $i$ in $\mathbf{g}$ . Also denotes the set of neighbors of $i$ .
$\mathcal{P}_{a_1 a_n}(\mathbf{g})$	a path from agent $a_1$ to $a_n$ in $\mathbf{g}$ such that $\langle a_1, a_2 \rangle, \langle a_2, a_3 \rangle, \dots, \langle a_{n-1}, a_n \rangle \in \mathcal{L}$ .
$d_{ij}(\mathbf{g})$	the length of the shortest path connecting agents $i$ and $j$ in $\mathbf{g}$ .
$\mathbf{g} + \langle ij \rangle$	new link $\langle ij \rangle$ is added to $\mathbf{g}$ .
$\mathbf{g} - \langle ij \rangle$	existing link $\langle ij \rangle$ is deleted from $\mathbf{g}$ .
$\mathcal{G}(N)$	the set of all networks on $N$ agents.
$u_i(\mathbf{g})$	utility of agent $i$ in $\mathbf{g}$ .

## 2.2. Storage Sharing

According to [1], agents could limit storage-sharing with those who are close to them in the social cloud. In order to capture this, we make use of the harmonic centrality measure (discussed in [10,27,28]), defined as follows:

$$\Phi_i(\mathbf{g}) = \sum_{j \in \mathbf{g} \setminus \{i\}} \frac{1}{d_{ij}(\mathbf{g})}. \quad (1)$$

We use harmonic centrality as it deals with disconnected networks as well.

In  $\mathbf{g}$ , an agent  $j$  (as a storage provider) computes a probability distribution on all agents for the purpose of allocating storage space to agent  $i \in \mathbf{g}$  (as a data owner), as below:

$$\alpha_{ij}(\mathbf{g}) = \frac{\frac{1}{d_{ij}(\mathbf{g})}}{\sum_{j \in \mathbf{g} \setminus \{i\}} \frac{1}{d_{ij}(\mathbf{g})}} = \frac{1}{d_{ij}(\mathbf{g})\Phi_i(\mathbf{g})}, \quad (2)$$

where  $\alpha_{ij}(\mathbf{g})$  is the probability that agent  $i$  will obtain storage space from agent  $j$  in  $\mathbf{g}$ .

**Remark 1.** If  $d_{ij}(\mathbf{g}) = \infty$ , then  $\alpha_{ij}(\mathbf{g}) = 0$  (and  $\alpha_{ji}(\mathbf{g}) = 0$ ). As agents  $i$  and  $j$  are disconnected in  $\mathbf{g}$ , their chances of obtaining storage space from each other is zero.

The probability that an agent  $i$  obtains storage space from at least one agent in  $\mathbf{g}$  is

$$\gamma_i(\mathbf{g}) = 1 - \prod_{j \in \mathbf{g} \setminus \{i\}} (1 - \alpha_{ij}(\mathbf{g})).$$

## 2.3. Agent's Utility and Symmetry

The utility of agent  $i$  in  $\mathbf{g}$  is given by a function  $u_i : \mathcal{G} \rightarrow \mathbb{R}^+$ . Let  $u$  be the the vector (profile) of utility functions  $u = (u_1, \dots, u_n)$ . Thus, we have  $u : \mathcal{G} \rightarrow \mathbb{R}^N$ . In other words, each possible social storage cloud structure  $\mathbf{g} \in \mathcal{G}$  leads to a utility function profile for agents.

We define the utility of agents in a social storage cloud  $\mathbf{g}$  with the following parameters. An agent,  $i$ , loses its data with probability  $\lambda_i \in (0, 1)$ . Therefore, to minimize this risk of data loss, agent  $i$  aims to backup its data on the storage provided by others. For agent  $i$ ,  $\beta_i$  is the value of the local data that is to be backed up. Agent  $i$  obtains storage space provided by others in  $\mathbf{g}$ , with probability  $\gamma_i(\mathbf{g})$ . Thus, the value of data  $\beta_i$ , the chance of losing the data  $\lambda_i$ , and the chance of obtaining storage space  $\gamma_i(\mathbf{g})$  capture the expected benefit of agent  $i$  in  $\mathbf{g}$ .

An agent searches for storage by staying connected in the network. Direct as well as indirect links help agents to get storage space. The direct link between agents  $i$  and  $j$  costs  $\zeta_i$ . This cost can be interpreted as the cost required for maintaining storage space, infrastructure, bandwidth, time, and so on. The cost to maintain an existing link and that for adding (and maintaining) a new link are the same. There is no *additional* cost to add a new link. Thus, agent  $i$  incurs a total cost of  $\zeta_i\eta_i(\mathbf{g})$  in order to obtain an expected benefit of  $\beta_i\lambda_i\gamma_i(\mathbf{g})$ , in case of data loss. But the network is formed upfront, before the data loss happens. The cost to maintain links is, hence, incurred even in the case of no data loss, where the expected benefit to  $i$  is  $\beta_i(1 - \lambda_i)$ .

Therefore, given the aforementioned parameters, the expected utility is

$$u_i(\mathbf{g}) = \beta_i(1 - \lambda_i) + \beta_i\lambda_i\gamma_i(\mathbf{g}) - \zeta_i\eta_i(\mathbf{g}). \quad (3)$$

Free riding (a situation where an agent offers less storage space, but consumes more) is a widely discussed issue in the literature on peer-to-peer storage. In order to deal with free riding, many P2P storage systems (for example, Internet Cooperative Backup System [29], PeerStore [30], Pastiche [31]) follow a symmetric storage-sharing mechanism, where agents share the same amount of storage space.

We define a symmetric social storage cloud  $\mathbf{g}$  as follows.

**Definition 1.** A symmetric social storage cloud (SSSC)  $\mathbf{g}$  is a network where the benefit (value) associated with backed-up data is the same for all agents in the network, that is,  $\beta_i = \beta_j$  (say  $\beta$ ),  $\zeta_i = \zeta_j$  (say  $\zeta^7$ ), and  $\lambda_i = \lambda_j$  (say  $\lambda^8$ ) for all  $i, j \in \mathcal{A}$ , and hence, utility of each agent  $i$  in  $\mathbf{g}$  is

$$u_i(\mathbf{g}) = \beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g}) - \zeta\eta_i(\mathbf{g}), \quad (4)$$

where  $\lambda, \beta, \zeta \in (0, 1)$ .

For further study, we consider the above utility function (Equation (4)). Henceforth, whenever we refer to a network, or just  $\mathbf{g}$ , we mean an SSSC.

#### 2.4. Pairwise Stability

In order to characterize endogenously built social storage cloud, we adopt *pairwise stability* [26] as a solution concept. A network is pairwise stable if (1) no agent benefits by deleting an existing link and (2) no two agents benefit by adding a new link between them.

**Definition 2.** [26] A network  $\mathbf{g}$  is pairwise stable if

- for all  $i, j \in \mathbf{g}$  such that  $\langle ij \rangle \in \mathbf{g}$ ,  $u_i(\mathbf{g}) \geq u_i(\mathbf{g} - \langle ij \rangle)$ , and  $u_j(\mathbf{g}) \geq u_j(\mathbf{g} - \langle ij \rangle)$ ; and
- for all  $i, j \in \mathbf{g}$  such that  $\langle ij \rangle \notin \mathbf{g}$ , if  $u_i(\mathbf{g} + \langle ij \rangle) > u_i(\mathbf{g})$ , then  $u_j(\mathbf{g} + \langle ij \rangle) < u_j(\mathbf{g})$ .

### 3. Network Structure and Storage Availability

One of the objectives of this paper is to understand the impact of link addition and deletion on storage availability for those agents who are involved in the link addition/deletion as well as those who are not. The storage availability is determined by the distances between them and their closeness (from Equation (2)). Therefore, first we study how addition and deletion of a link impacts the shortest distances between pairs of agents and, therefore, their closeness. This analysis provides a base for understanding the effect of link-addition/deletion on agents' storage availability in  $\mathbf{g}$ .

#### 3.1. Effect of Link Alteration on Closeness

**Lemma 1.** Suppose  $\langle ij \rangle \notin \mathbf{g}$ . Then,  $\Phi_i(\mathbf{g} + \langle ij \rangle) > \Phi_i(\mathbf{g})$ .

**Proof.** Clearly,  $d_{ij}(\mathbf{g} + \langle ij \rangle) < d_{ij}(\mathbf{g})$ . As  $\langle ij \rangle \notin \mathbf{g}$ , we have,  $d_{ij}(\mathbf{g}) \geq 2$ . Also,  $d_{ij}(\mathbf{g} + \langle ij \rangle) = 1$ . Thus,  $\Phi_i(\mathbf{g})$  and  $\Phi_j(\mathbf{g})$  increase by at least  $\frac{d_{ij}(\mathbf{g})-1}{d_{ij}(\mathbf{g})}$  in  $\mathbf{g} + \langle ij \rangle$ .  $\square$

**Lemma 2.** Suppose  $\langle ij \rangle \in \mathbf{g}$ . Then,  $\Phi_i(\mathbf{g} - \langle ij \rangle) < \Phi_i(\mathbf{g})$ .

**Proof.**

- Let us assume there is no path between  $i$  and  $j$  in  $\mathbf{g} - \langle ij \rangle$ , then  $d_{ij}(\mathbf{g} - \langle ij \rangle) = \infty$ , thus,  $\Phi_i(\mathbf{g})$  and  $\Phi_j(\mathbf{g})$  decrease by 1 in  $\mathbf{g} - \langle ij \rangle$ .
- Now, let us assume there exists a path  $\mathcal{P}_{ij}(\mathbf{g} - \langle ij \rangle)$  between  $i$  and  $j$  in  $\mathbf{g} - \langle ij \rangle$ , the distance between  $i$  and  $j$  in  $\mathbf{g} - \langle ij \rangle$  being at least 1 more than that in  $\mathbf{g}$ . Thus,  $\Phi_i(\mathbf{g})$  and  $\Phi_j(\mathbf{g})$  decrease by at least  $\frac{d_{ij}(\mathbf{g}-\langle ij \rangle)-1}{d_{ij}(\mathbf{g}-\langle ij \rangle)}$  in  $\mathbf{g} - \langle ij \rangle$ .  $\square$

<sup>7</sup> We assume,  $\zeta = \frac{\zeta_i + \zeta_j}{2}$ , that is, a pair of agents involved in a link share the cost  $\zeta$ .

<sup>8</sup> For simplicity, we assume uniform data loss rate  $\lambda$ .

Lemmas 1 and 2 show that, with respect to closeness, every link benefits agents on either side of the link. An action of link addition or deletion between a pair of agents not only impacts their closeness, but also that of other agents. Now, we study the impact of link addition or deletion between a pair of agents (say,  $i$  and  $j$ ) on the closeness of the other agents  $k \in \mathfrak{g} \setminus \{i, j\}$ .

**Lemma 3.** *Suppose  $\langle ij \rangle \notin \mathfrak{g}$  and  $k \in \mathfrak{g} \setminus \{i, j\}$ . Then,  $\Phi_k(\mathfrak{g}) = \Phi_k(\mathfrak{g} + \langle ij \rangle)$  if and only if  $d_{kl}(\mathfrak{g}) = d_{kl}(\mathfrak{g} + \langle ij \rangle)$  for all  $l \in \mathfrak{g}$ .*

**Proof.** If  $d_{kl}(\mathfrak{g}) = d_{kl}(\mathfrak{g} + \langle ij \rangle)$  for all  $l \in \mathfrak{g}$ , then by Equation (1),  $\Phi_k(\mathfrak{g}) = \Phi_k(\mathfrak{g} + \langle ij \rangle)$ .

Conversely, suppose  $\Phi_k(\mathfrak{g}) = \Phi_k(\mathfrak{g} + \langle ij \rangle)$ .

It is easy to see that, if for some  $l \in \mathfrak{g}$ , if  $d_{kl}(\mathfrak{g} + \langle ij \rangle) \neq d_{kl}(\mathfrak{g})$ , then  $d_{kl}(\mathfrak{g} + \langle ij \rangle) < d_{kl}(\mathfrak{g})$ . (Paths in  $\mathfrak{g}$  exist in  $\mathfrak{g} + \langle ij \rangle$  too).

We have  $d_{kl}(\mathfrak{g} + \langle ij \rangle) \leq d_{kl}(\mathfrak{g})$  for all  $l \in \mathfrak{g}$  and, if there exists  $x$  such that  $d_{kx}(\mathfrak{g} + \langle ij \rangle) < d_{kx}(\mathfrak{g})$ , then  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ , a contradiction.  $\square$

**Lemma 4.** *Suppose  $\langle ij \rangle \in \mathfrak{g}$  and  $k \in \mathfrak{g} \setminus \{i, j\}$ . Then,  $\Phi_k(\mathfrak{g}) = \Phi_k(\mathfrak{g} - \langle ij \rangle)$  if and only if  $d_{kl}(\mathfrak{g}) = d_{kl}(\mathfrak{g} - \langle ij \rangle)$  for all  $l \in \mathfrak{g}$ .*

**Proof.** As  $d_{kl}(\mathfrak{g} - \langle ij \rangle) \geq d_{kl}(\mathfrak{g})$  for all  $l$ , the proof follows in lines similar to that of Lemma 3.  $\square$

We now show necessary and sufficient conditions for increase in the closeness of agents who are not involved in link addition or deletion.

**Theorem 1.** *Suppose  $\langle ij \rangle \notin \mathfrak{g}$ , and let  $k$  be an agent distinct from  $i$  and  $j$ . Then,  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$  if and only if there exists at least one agent  $l \in \mathfrak{g}$  such that  $d_{kl}(\mathfrak{g}) \geq 3$  and all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g} + \langle ij \rangle)$  from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  contain  $\langle ij \rangle$ .*

**Proof.** Let  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ . Then, by Lemma 3, there must be at least one agent, say  $l$ , such that  $d_{kl}(\mathfrak{g}) > d_{kl}(\mathfrak{g} + \langle ij \rangle)$ .

Suppose  $i, k$ , and  $l$  are all distinct. Note that  $j$  may be the same as  $l$ .

If possible, let  $d_{kl}(\mathfrak{g}) < d_{ki}(\mathfrak{g}) + d_{ij}(\mathfrak{g}) + d_{jl}(\mathfrak{g})$  for all  $l \in \mathfrak{g}$ . Then,  $d_{kl}(\mathfrak{g}) = d_{kl}(\mathfrak{g} + \langle ij \rangle)$  for all  $l \in \mathfrak{g}$ . From Lemma 1,  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ , a contradiction. Therefore, there exists an  $l \in \mathfrak{g}$  such that  $d_{kl}(\mathfrak{g}) = d_{ki}(\mathfrak{g}) + d_{ij}(\mathfrak{g}) + d_{jl}(\mathfrak{g})$ .

As  $\langle ij \rangle \notin \mathfrak{g}$ ,  $d_{ij}(\mathfrak{g}) \geq 2$ . As  $k \neq i$ ,  $d_{ik}(\mathfrak{g}) \geq 1$  and  $j = l$ . Hence,  $d_{kl}(\mathfrak{g}) \geq 3$ .

$$\begin{aligned} \text{Now, } d_{kl}(\mathfrak{g} + \langle ij \rangle) &= d_{ki}(\mathfrak{g} + \langle ij \rangle) + d_{ij}(\mathfrak{g} + \langle ij \rangle) + d_{jl}(\mathfrak{g} + \langle ij \rangle) \\ &= d_{ki}(\mathfrak{g}) + d_{ij}(\mathfrak{g} + \langle ij \rangle) + d_{jl}(\mathfrak{g}) \\ &< d_{ki}(\mathfrak{g}) + d_{ij}(\mathfrak{g}) + d_{jl}(\mathfrak{g}) \\ &= d_{kl}(\mathfrak{g}). \end{aligned}$$

It follows that every shortest path between  $k$  and  $l$  in  $\mathfrak{g} + \langle ij \rangle$  contains  $\langle ij \rangle$ . (Note that if there exists a shortest path from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  that does not contain  $\langle ij \rangle$ , then this shortest path exists in  $\mathfrak{g}$  too).

Conversely, let  $l \in \mathfrak{g}$  such that  $d_{kl}(\mathfrak{g}) \geq 3$  and all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g} + \langle ij \rangle)$  from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  contain  $\langle ij \rangle$ .

Clearly,  $\Phi_k(\mathfrak{g}) \leq \Phi_k(\mathfrak{g} + \langle ij \rangle)$ .

If possible, let  $\Phi_k(\mathfrak{g}) = \Phi_k(\mathfrak{g} + \langle ij \rangle)$ . This means for every  $l$  in  $\mathfrak{g}$  there exists a shortest path from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  that does not contain  $\langle ij \rangle$ , a contradiction. Therefore,  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ .  $\square$

**Theorem 2.** *Suppose  $\langle ij \rangle \in \mathfrak{g}$ , and let  $k$  be an agent distinct from  $i$  and  $j$ . Then,  $\Phi_k(\mathfrak{g} - \langle ij \rangle) < \Phi_k(\mathfrak{g})$  if and only if there exists at least one agent  $l \in \mathfrak{g}$  such that  $d_{kl}(\mathfrak{g}) \geq 2$  and all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g})$  from  $k$  to  $l$  in  $\mathfrak{g}$  contain  $\langle ij \rangle$ .*

We skip the proof as it is similar to the proof of Theorem 1.



In subsequent sections, we present our results due to link addition. We present our results on link deletion in Appendix B.

### 3.2. Effect of Closeness on Distances of Agents Not Involved in Link Alteration

In this section, we classify agents whose mutual distances from each other remain the same after link alteration. We use the same to analyze the effect of closeness on distances between agents who are not involved in the link addition or deletion.

Given  $k$  such that  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ , we use  $L_k^+$  to denote the set of all  $l \in \mathfrak{g}$  such that all shortest paths from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  contain  $\langle ij \rangle$ . We use  $l_k^+$  to denote an agent in  $L_k^+$ .

**Proposition 1.** *Suppose  $i, j$ , and  $k$  are distinct agents in  $\mathfrak{g}$ . Suppose  $l$  is another agent, distinct from  $i$  and  $k$ , and suppose  $\Phi_k(\mathfrak{g} + \langle ij \rangle) > \Phi_k(\mathfrak{g})$ . If  $d_{ki}(\mathfrak{g} + \langle ij \rangle) < d_{kj}(\mathfrak{g} + \langle ij \rangle) \leq d_{kl}(\mathfrak{g} + \langle ij \rangle)$ , then  $d_{ik}(\mathfrak{g} + \langle ij \rangle) = d_{ik}(\mathfrak{g})$ .*

**Proof.** We have  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ . Then, from Theorem 1, there exists  $l \in \mathfrak{g}$  such that all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g} + \langle ij \rangle)$  from  $k$  to  $l$  in  $\mathfrak{g} + \langle ij \rangle$  contain  $\langle ij \rangle$ .

We consider the two cases  $j = l$  and  $j \neq l$ .

1. Suppose  $j = l$ . As  $d_{ki}(\mathfrak{g} + \langle ij \rangle) < d_{kj}(\mathfrak{g} + \langle ij \rangle)$ ,  $k$  observes  $i$  before  $j$  on all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g} + \langle ij \rangle)$ . This implies  $d_{ik}(\mathfrak{g} + \langle ij \rangle) = d_{ik}(\mathfrak{g})$ .
2. Suppose  $j \neq l$ . As  $d_{ki}(\mathfrak{g} + \langle ij \rangle) < d_{kj}(\mathfrak{g} + \langle ij \rangle) \leq d_{kl}(\mathfrak{g} + \langle ij \rangle)$ ,  $k$  observes  $i$  before  $j$ , and  $j$  before  $l$ , on all shortest paths  $\mathcal{P}_{kl}(\mathfrak{g} + \langle ij \rangle)$ . This implies  $d_{ik}(\mathfrak{g} + \langle ij \rangle) = d_{ik}(\mathfrak{g})$ .  $\square$

**Definition 3.** *Suppose  $\langle ij \rangle \notin \mathfrak{g}$  and  $k$  is an agent such that  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$ . A  $(k, +ij)$ -shortest-path-network,  $\mathfrak{g}_{ij}^{k+}$ , is a subnetwork of  $\mathfrak{g} + \langle ij \rangle$  that consists of all shortest paths from  $k$  to  $l_k^+$  in  $\mathfrak{g} + \langle ij \rangle$ , which contain  $\langle ij \rangle$ , for all  $l_k^+ \in L_k^+$ .*

**Definition 4.** *An (all  $k, +ij$ )-shortest-path-network,  $\mathfrak{g}_i^+$ , is  $\bigcup_{\substack{k \in \mathfrak{g}, \\ \Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)}} \mathfrak{g}_{ij}^{k+}$ , the smallest network consisting of all  $(k, +ij)$ -shortest-path-networks.*

**Definition 5.** *A sub- $(i, +)$ -network,  $\mathfrak{g}_i^+$  of  $\mathfrak{g}_i^+$ , is the induced subnetwork of  $\mathfrak{g}_i^+$  consisting of all agents  $k \in \mathfrak{g}_{ij}^+$  such that  $d_{ik}(\mathfrak{g}) = d_{ik}(\mathfrak{g} + \langle ij \rangle)$ . Similarly, we define the sub- $(j, +)$ -network,  $\mathfrak{g}_j^+$  of  $\mathfrak{g}_j^+$ , as the induced subnetwork of  $\mathfrak{g}_j^+$  consisting of all agents  $l \in \mathfrak{g}_{ij}^+$  such that  $d_{jl}(\mathfrak{g}) = d_{jl}(\mathfrak{g} + \langle ij \rangle)$ .*

Refer to Appendix A for an illustration of the above definitions.

**Proposition 2.** *For all  $k, \bar{k} \in \mathfrak{g}_i^+$ ,  $d_{k\bar{k}}(\mathfrak{g}) = d_{k\bar{k}}(\mathfrak{g} + \langle ij \rangle)$ .*

**Proof.** If  $k, \bar{k} \in \mathfrak{g}_i^+$  then, from Definition 5,  $d_{ik}(\mathfrak{g}) = d_{ik}(\mathfrak{g} + \langle ij \rangle)$  and  $d_{i\bar{k}}(\mathfrak{g}) = d_{i\bar{k}}(\mathfrak{g} + \langle ij \rangle)$ . As  $k, \bar{k} \in \mathfrak{g}_{ij}^+$  as well, there exists  $l$  and  $\bar{l}$  such that  $d_{kl}(\mathfrak{g}) > d_{kl}(\mathfrak{g} + \langle ij \rangle)$  and  $d_{\bar{k}\bar{l}}(\mathfrak{g}) > d_{\bar{k}\bar{l}}(\mathfrak{g} + \langle ij \rangle)$ .

It is sufficient to show that, given  $\bar{k}, \bar{l}$  can never be  $k$ .

If possible, let  $\bar{l} = k$ . Then, from Definition 5,  $d_{\bar{k}i}(\mathfrak{g}) = d_{\bar{k}i}(\mathfrak{g} + \langle ij \rangle)$  implies  $\bar{k}$  observes  $i$  first, and subsequently  $j$  to reach  $k$ , on all shortest paths  $\mathcal{P}_{\bar{k}k}(\mathfrak{g} + \langle ij \rangle)$  from  $\bar{k}$  to  $k$  in  $\mathfrak{g} + \langle ij \rangle$ . Then,  $d_{ik}(\mathfrak{g}) \neq d_{ik}(\mathfrak{g} + \langle ij \rangle)$ . This is because, if  $k = j$ ,  $d_{ik}(\mathfrak{g}) < d_{ik}(\mathfrak{g} + \langle ij \rangle) = 1$ . Therefore,  $k \notin \mathfrak{g}_i^+$ , which is a contradiction. Now, if  $k \neq j$ , then  $k$  must first visit  $j$ , and later  $i$ , to reach  $\bar{k}$  on all shortest paths  $\mathcal{P}_{\bar{k}k}(\mathfrak{g} + \langle ij \rangle)$  from  $\bar{k}$  to  $k$ . This implies  $d_{ik}(\mathfrak{g}) \neq d_{ik}(\mathfrak{g} + \langle ij \rangle)$  and hence,  $k \notin \mathfrak{g}_i^+$ , again, a contradiction. Thus,  $k \neq \bar{l}$ .  $\square$

We discuss our results on shortest distances due to link deletion in Appendix B.1.

### 3.3. Effect of Link Alteration on Storage Availability

Our aim here is to analyze under what conditions agents' chance of obtaining storage space in the network increases or decreases by adding a new link. We present our results in the case of link deletion in Appendix B.2.

**Lemma 5.** Suppose agent  $i$  and  $j$  add a direct link in  $\mathbf{g}$  and let  $k \notin \mathbf{g}_{ij}^+$ . Then,  $\alpha_{ik}(\mathbf{g}) = \alpha_{ik}(\mathbf{g} + \langle ij \rangle)$  and  $\alpha_{jk}(\mathbf{g}) = \alpha_{jk}(\mathbf{g} + \langle ij \rangle)$ .

**Proof.** If agent  $k \notin \mathbf{g}_{ij}^+$  then  $\Phi_k(\mathbf{g}) = \Phi_k(\mathbf{g} + \langle ij \rangle)$ . Thus,  $d_{ki}(\mathbf{g}) = d_{ki}(\mathbf{g} + \langle ij \rangle)$ . Therefore, from Equation (2),  $\alpha_{ik}(\mathbf{g}) = \alpha_{ik}(\mathbf{g} + \langle ij \rangle)$ . A similar proof holds for  $j$  too.  $\square$

**Lemma 6.** Suppose agents  $i, j, k$ , and  $l$  are such that  $i \neq j, j \neq k, i \neq l$ , and  $k \neq l$ . (Agents  $i$  and  $k$  may be the same, and agents  $j$  and  $l$  may be the same). Suppose  $\langle ij \rangle \notin \mathbf{g}, k \in \mathbf{g}_i^+,$  and  $l \in \mathbf{g}_j^+$ . Then,

1.  $\alpha_{kl}(\mathbf{g}) < \alpha_{kl}(\mathbf{g} + \langle ij \rangle)$ , and
2.  $i \neq k$  implies that  $\alpha_{ik}(\mathbf{g}) > \alpha_{ik}(\mathbf{g} + \langle ij \rangle)$ . Similarly, if  $j \neq l$ , then  $\alpha_{jl}(\mathbf{g}) > \alpha_{jl}(\mathbf{g} + \langle ij \rangle)$ .

**Proof.** Refer Appendix C for the proof.  $\square$

**Lemma 7.** Let  $k$  and  $\bar{k}$  be agents in  $\mathbf{g}_i^+$ . Then,  $\alpha_{k\bar{k}}(\mathbf{g}) = \alpha_{k\bar{k}}(\mathbf{g} + \langle ij \rangle)$  and  $\alpha_{\bar{k}k}(\mathbf{g}) = \alpha_{\bar{k}k}(\mathbf{g} + \langle ij \rangle)$ .

**Proof.** The proof follows from Proposition 2.  $\square$

**Theorem 3.** Suppose agents  $i$  and  $j$  are such  $i \neq j$ , and  $\langle ij \rangle \notin \mathbf{g}$ . Then,  $\gamma_i(\mathbf{g}) < \gamma_i(\mathbf{g} + \langle ij \rangle)$  if and only

$$\text{if } \frac{\prod_{k \in \mathbf{g}_i^+} (1 - \alpha_{ik}(\mathbf{g} + \langle ij \rangle))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{il}(\mathbf{g}))} < \frac{\prod_{k \in \mathbf{g}_i^+} (1 - \alpha_{ik}(\mathbf{g}))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{il}(\mathbf{g} + \langle ij \rangle))}.$$

$$\text{Additionally, } \gamma_i(\mathbf{g}) < \gamma_i(\mathbf{g} + \langle ij \rangle) \text{ if and only if } \frac{\prod_{k \in \mathbf{g}_i^+} (\alpha_{ik}(\mathbf{g} + \langle ij \rangle))}{\prod_{l \in \mathbf{g}_j^+} (\alpha_{il}(\mathbf{g}))} > \frac{\prod_{k \in \mathbf{g}_i^+} (\alpha_{ik}(\mathbf{g}))}{\prod_{l \in \mathbf{g}_j^+} (\alpha_{il}(\mathbf{g} + \langle ij \rangle))}.$$

**Proof.** The proof follows from Lemmas 5, 6, and 7.  $\square$

### 3.4. Externalities

In this section, we study externalities, that is, how a link that is added between a pair of agents affects the utility of others. (Refer to Definition 6). The particular form of externalities (positive, negative, or none) is crucial in determining which network is likely to evolve and the conditions under which it will lead to a stable and efficient network.

**Definition 6.** [32] Consider a network,  $\mathbf{g}$ , with agents  $i, j \in \mathbf{g}$  such that  $i \neq j$  and  $\langle ij \rangle \notin \mathbf{g}$ . Suppose agents  $i$  and  $j$  form a direct link  $\langle ij \rangle$ . Then, agent  $k \in \mathbf{g} \setminus \{i, j\}$  experiences

1. Positive externalities if  $u_k(\mathbf{g} + \langle ij \rangle) > u_k(\mathbf{g})$ ;
2. Negative externalities if  $u_k(\mathbf{g} + \langle ij \rangle) < u_k(\mathbf{g})$ ;
3. No externalities if  $u_k(\mathbf{g} + \langle ij \rangle) = u_k(\mathbf{g})$ .

We now show that the type of externalities an agent  $k \in \mathbf{g}$  experiences, can be determined using conditions on the storage availability, independent of the data loss rate and the value that agents associate with their data.

**Proposition 3.** In an SSSC  $\mathbf{g}$ , an agent  $k \in \mathbf{g}$  experiences

1. Positive externalities if  $\gamma_k(\mathbf{g} + \langle ij \rangle) > \gamma_k(\mathbf{g})$ ;



2. Negative externalities if  $\gamma_k(\mathbf{g} + \langle ij \rangle) < \gamma_k(\mathbf{g})$ ;
3. No externalities if  $\gamma_k(\mathbf{g} + \langle ij \rangle) = \gamma_k(\mathbf{g})$ .

**Proof.**

1. By Definition 6,  $u_k(\mathbf{g} + \langle ij \rangle) > u_k(\mathbf{g})$   
 $\Rightarrow \beta(1 - \lambda) + \beta\lambda\gamma_k(\mathbf{g} + \langle ij \rangle) - \zeta\eta_k(\mathbf{g} + \langle ij \rangle) > \beta(1 - \lambda) + \beta\lambda\gamma_k(\mathbf{g}) - \zeta\eta_k(\mathbf{g})$ .  
 As agent  $k$  does not pay the cost for link  $\langle ij \rangle$ , we have  $\zeta\eta_k(\mathbf{g} + \langle ij \rangle) = \zeta\eta_k(\mathbf{g})$ .  
 Thus,  $\beta\lambda\gamma_k(\mathbf{g} + \langle ij \rangle) > \beta\lambda\gamma_k(\mathbf{g}) \Rightarrow \gamma_k(\mathbf{g} + \langle ij \rangle) > \gamma_k(\mathbf{g})$ .
2. For Cases 2 and 3, the proof is similar to that of Case 1.  $\square$

The following results provide a necessary and sufficient condition under which an agent  $k \in \mathbf{g}$  experiences positive or negative externalities.

**Lemma 8.** Let  $i, j$ , and  $k$  be distinct agents in  $\mathbf{g}$ . Suppose  $k \notin \mathbf{g}_{ij}^+$ . Then,  $k$  experiences only negative externalities.

**Proof.** If agents  $i$  and  $j$  add a direct link in  $\mathbf{g}$ , then, from Lemma 1,  $\Phi_i(\mathbf{g}) < \Phi_i(\mathbf{g} + \langle ij \rangle)$ . If  $k \notin \mathbf{g}_{ij}^+$ , then, from Theorem 1,  $\Phi_k(\mathbf{g}) = \Phi_k(\mathbf{g} + \langle ij \rangle)$ , thus,  $d_{kl}(\mathbf{g}) = d_{kl}(\mathbf{g} + \langle ij \rangle)$  for all  $l \in \mathbf{g}$ . Therefore,  $\alpha_{ki}(\mathbf{g} + \langle ij \rangle) < \alpha_{ki}(\mathbf{g})$ , by Equation (2). Now, for all  $l \in \mathbf{g}$ ,  $\Phi_l(\mathbf{g}) \leq \Phi_l(\mathbf{g} + \langle ij \rangle)$ . If  $\Phi_l(\mathbf{g}) = \Phi_l(\mathbf{g} + \langle ij \rangle)$ , then  $\alpha_{kl}(\mathbf{g} + \langle ij \rangle) = \alpha_{kl}(\mathbf{g})$  and, if  $\Phi_l(\mathbf{g}) < \Phi_l(\mathbf{g} + \langle ij \rangle)$ , then  $\alpha_{kl}(\mathbf{g} + \langle ij \rangle) < \alpha_{kl}(\mathbf{g})$ . Thus,  $\gamma_k(\mathbf{g} + \langle ij \rangle) < \gamma_k(\mathbf{g})$ .  $\square$

**Theorem 4.** Suppose agents  $i, j, k, \bar{k}$ , and  $l$  are such that  $i \neq j, i \neq k, i \neq l, j \neq k, k \neq \bar{k}$ , and  $k \neq l$ . (Agents  $i$  and  $\bar{k}$  may be the same, and agents  $j$  and  $l$  may be the same). Suppose  $\langle ij \rangle \notin \mathbf{g}, \bar{k} \in \mathbf{g}_i^+$  and  $l \in \mathbf{g}_j^+$ . Then, agent  $k$  experiences positive externalities if and only if

$$k \in \mathbf{g}_{ij}^+ \text{ and } \frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g} + \langle ij \rangle))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g}))} < \frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g}))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g} + \langle ij \rangle))}, \text{ otherwise } k \text{ experiences negative externalities.}$$

**Proof.** From Lemma 8, it is required to increment in agent  $k$ 's closeness. It is straightforward to observe that  $\frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g} + \langle ij \rangle))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g}))} < \frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g}))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g} + \langle ij \rangle))}$ , then  $\gamma_k(\mathbf{g} + \langle ij \rangle) > \gamma_k(\mathbf{g})$ . Thus,  $k$  experiences positive externalities.

Conversely, let  $\gamma_k(\mathbf{g} + \langle ij \rangle) < \gamma_k(\mathbf{g})$ , then either from Proposition 8,  $d_{ki}(\mathbf{g}) = d_{ki}(\mathbf{g} + \langle ij \rangle)$ , for all  $i \in \mathbf{g}$  or  $\frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g} + \langle ij \rangle))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g}))} < \frac{\prod_{\bar{k} \in \mathbf{g}_i^+} (1 - \alpha_{k\bar{k}}(\mathbf{g}))}{\prod_{l \in \mathbf{g}_j^+} (1 - \alpha_{kl}(\mathbf{g} + \langle ij \rangle))}$ .  $\square$

Lemma 8 and Theorem 4 show that an increase in the closeness of an agent (who is not involved in the link formation) is necessary in order for that agent to experience positive externalities. Although we have provided a necessary and sufficient condition for positive and negative externalities by performing a microscopic analysis of externalities, it is hard to obtain a general characterization of networks where agents experience only positive externalities. This leads us to the following question. At least for specific network structures, can we show positive (or negative) externalities? For instance, we can argue that in a two diameter network, agents never experience positive externalities.

#### 4. Characterization of Stable and Efficient Networks

One of the central focuses of this study is to analyze what network is likely to emerge when each agent (or pair of agents) decides selfishly which link they want to delete (respectively, whether to add

a link or not), when agents build their social connections (links) based on the benefit associated with their data, the cost for link formation, and the prevailing data loss rate.

In the following subsections, we discuss pairwise stable networks, efficient networks, and the measures of efficiency, namely, price of anarchy (PoA) and price of stability (PoS). In our analysis of stable and efficient networks, we assume that network formation takes place starting with the null network (where there are no links between any pair of agents).

#### 4.1. Stable Networks: Characterization, Existence, and Uniqueness

We now discuss the conditions under which an agent prefers to add a new link or delete an existing link, and use the same to characterize stable networks.

**Lemma 9.** Let  $\langle ij \rangle \notin \mathbf{g}$ . An agent  $i \in \mathbf{g}$  is benefited by adding a direct link with agent  $j \in \mathbf{g}$  if and only if  $\beta\lambda[\gamma_i(\mathbf{g} + \langle ij \rangle) - \gamma_i(\mathbf{g})] > \varsigma$ .

**Proof.** Agent  $i$  has incentive to form a link with agent  $j$  if and only if  $u_i(\mathbf{g} + \langle ij \rangle) > u_i(\mathbf{g})$   
 $\Rightarrow \beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g} + \langle ij \rangle) - \varsigma(\eta_i(\mathbf{g}) + 1) > \beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g}) - \varsigma\eta_i(\mathbf{g})$   
 $\Rightarrow \beta\lambda[\gamma_i(\mathbf{g} + \langle ij \rangle) - \gamma_i(\mathbf{g})] > \varsigma$ .  $\square$

**Corollary 1.** An agent  $i \in \mathbf{g}$  has no incentive to add a link with agent  $j \in \mathbf{g}$  if and only if  $\lambda[\gamma_i(\mathbf{g} + \langle ij \rangle) - \gamma_i(\mathbf{g})] \leq \frac{\varsigma}{\beta}$ .

**Lemma 10.** Let  $\langle ij \rangle \in \mathbf{g}$ . An agent  $i \in \mathbf{g}$  benefits by deleting a link with agent  $j$  if and only if  $\beta\lambda[\gamma_i(\mathbf{g}) - \gamma_i(\mathbf{g} - \langle ij \rangle)] < \varsigma$ .

**Proof.** An agent  $i$  has incentive to delete a link with agent  $j$  if and only if  $u_i(\mathbf{g} - \langle ij \rangle) > u_i(\mathbf{g})$ .  
 $\Rightarrow \beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g} - \langle ij \rangle) - \varsigma(\eta_i(\mathbf{g}) - 1) > \beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g}) - \varsigma\eta_i(\mathbf{g})$   
 $\Rightarrow \varsigma > \beta\lambda[\gamma_i(\mathbf{g}) - \gamma_i(\mathbf{g} - \langle ij \rangle)]$ .  $\square$

**Corollary 2.** An agent  $i$  has no incentive to delete an existing link with agent  $j$  if and only if  $\lambda[\gamma_i(\mathbf{g}) - \gamma_i(\mathbf{g} - \langle ij \rangle)] \geq \frac{\varsigma}{\beta}$ .

Theorem 4, stated below, provides an easy characterization of a stable network  $\mathbf{g}$ .

**Proposition 4.** A network  $\mathbf{g}$  is pairwise stable if and only if

1. for all  $i, j \in \mathbf{g}$ ,  $\lambda[\gamma_i(\mathbf{g}) - \gamma_i(\mathbf{g} - \langle ij \rangle)] \geq \frac{\varsigma}{\beta}$ , and  $\lambda[\gamma_j(\mathbf{g}) - \gamma_j(\mathbf{g} - \langle ij \rangle)] \geq \frac{\varsigma}{\beta}$ ; and
2. for all  $i, j \in \mathbf{g}$ , if  $\lambda[\gamma_i(\mathbf{g} + \langle ij \rangle) - \gamma_i(\mathbf{g})] > \frac{\varsigma}{\beta}$ , then  $\lambda[\gamma_j(\mathbf{g} + \langle ij \rangle) - \gamma_j(\mathbf{g})] < \frac{\varsigma}{\beta}$ .

**Proof.** The proof follows from Definition 2, Corollary 2, and Lemma 9.  $\square$

In the following theorem, we prove existence and uniqueness of pairwise stable networks, given the values of the parameters.

**Theorem 5.** There always exists a pairwise stable network. Given  $N$ , there exists exactly two pairwise stable networks. For each  $\beta, \varsigma$ , and  $\lambda$ , the pairwise stable network  $\mathbf{g}$  is unique.

1. If  $\lambda \leq \frac{\varsigma}{\beta}$ , then  $\mathbf{g}$  is the null network.
2. If  $\lambda > \frac{\varsigma}{\beta}$ , then  $\mathbf{g}$  consists of
  - (a) a set of  $\frac{N}{2}$  connected pairs of agents, if  $N$  is even; or
  - (b) a set of  $\frac{N-1}{2}$  connected pairs of agents and one isolated agent, if  $N$  is odd.

**Proof.** Initially, all agents are isolated in  $\mathbf{g}$ , hence, for all  $i \in \mathbf{g}$ ,  $\gamma_i(\mathbf{g}) = 0$ .

If agents  $i$  and  $j$  form a direct link  $\langle ij \rangle$ , then  $\Phi_i(\mathbf{g} + \langle ij \rangle) = \Phi_j(\mathbf{g} + \langle ij \rangle) = 1$ .

Thus,  $\gamma_i(\mathbf{g} + \langle ij \rangle) = \gamma_j(\mathbf{g} + \langle ij \rangle) = 1$ .

However, from Lemma 9, agents  $i$  and  $j$  benefit by forming a direct link if and only if

$$\lambda\beta[\gamma_i(\mathbf{g} + \langle ij \rangle) - \gamma_i(\mathbf{g})] > \zeta \text{ and } \lambda\beta[\gamma_j(\mathbf{g} + \langle ij \rangle) - \gamma_j(\mathbf{g})] > \zeta, \text{ respectively.}$$

This implies that a pair of agents have no incentive to add a direct link if and only if  $\lambda \leq \frac{\zeta}{\beta}$ .

Therefore,  $\mathbf{g}$  is the null network. This completes the proof of 1.

Now, if  $\lambda > \frac{\zeta}{\beta}$ , then every pair of agents has an incentive to add a direct link. Suppose agents  $i$  and  $j$  add a direct link, and suppose link  $\langle ij \rangle$  is the only link in the network, say  $\mathbf{g}'$ . Let  $k$  be another agent, different from  $i$  and  $j$ , in  $\mathbf{g}'$ . Then,  $\gamma_i(\mathbf{g}' + \langle ik \rangle) = 1 - (1 - \frac{1}{15})^2$ .

By Lemma 9, agent  $i$  benefits by adding the link  $\langle ik \rangle$  if and only if  $\lambda[\gamma_i(\mathbf{g}' + \langle ik \rangle) - \gamma_i(\mathbf{g}')] > \frac{\zeta}{\beta}$ .

Here,  $\lambda[\gamma_i(\mathbf{g}' + \langle ik \rangle) - \gamma_i(\mathbf{g}')] < 0 \not> \frac{\zeta}{\beta}$ .

This implies that no agent benefits by adding more than one link, proving 2.  $\square$

#### 4.2. Efficient Network, Price of Anarchy, and Price of Stability

We analyze whether the network formed by self-interested agents is also efficient, that is, socially optimal or in other words, “good” for all the agents put together.

**Definition 7.** A social storage network  $\mathbf{g}$  is efficient with respect to utility profile  $(u_1, \dots, u_N)$  if

$$\sum_{i \in N} [\beta(1 - \lambda) + \beta\lambda\gamma_i(\mathbf{g}) - \zeta\eta_i(\mathbf{g})] \geq \sum_{i \in N} [\beta(1 - \lambda) + \beta\lambda\gamma_i(\bar{\mathbf{g}}) - \zeta\eta_i(\bar{\mathbf{g}})], \forall \bar{\mathbf{g}} \in \mathcal{G}(N).$$

It might be possible that when self-interested agents build their social connections for their own benefit, the resulting network formation will lead to a “bad” outcome from a societal viewpoint. That is, the resulting network may be advantageous for a set of agents, while other agents may not be benefited by the outcome. This results in an inefficient network. In this state of affairs, we would like to measure how far a pairwise stable network is from an efficient network. For this, we make use of the widely discussed measures, namely, price of anarchy (PoA) and price of stability (PoS). We define these measures as follows.

**Definition 8.** The price of anarchy (PoA) is the ratio of the worst sum of the utility value of an equilibrium network and the optimal sum of the utility value in any network.

**Definition 9.** The price of stability (PoS) is the ratio of the best sum of the utility value of an equilibrium network and the optimal sum of the utility value in any network.

**Theorem 6.** Every pairwise stable network is efficient. Therefore,  $PoA = 1$ . In addition, every efficient network is pairwise stable. Hence,  $PoS = 1$ .

**Proof.** The proof follows from Theorem 5 and Definition 7, the fact that network formation starts with the null network, and the fact ([32]) that  $PoS = 1$  if and only if every efficient network is pairwise stable, and  $PoA = 1$  if and only if all pairwise stable networks are efficient.  $\square$

In Appendix D, we discuss our experimental results on random stable networks where, for 150 random scenarios, no agent loses its data. That is, even if the disk of an agent fails, in our random experiments, the disk of the agent’s neighbor (from whom it can retrieve its data) is intact.

### 5. Conclusions

In this paper, we present the model of social storage cloud network formation, where agents (involved in storage sharing and data backup) wish to form a network strategically. The agents in this network strive for increasing the probability of obtaining storage space by minimizing the distances

with others. We propose a degree-distance-based utility function and use the same to study network formation. We also study the impact of the decision of link addition (deletion) between a pair of agents on shortest distances, closeness, and storage availability.

We study the deviation conditions under which agents have an incentive to add or delete a link in a given network structure. With these conditions, we analyze pairwise stability and efficiency of social storage cloud. We show that there always exist a unique pairwise stable network, which is also efficient. Hence, the price of anarchy and the price of stability are, both, one.

### 5.1. Research Implications

In the social cloud literature, the issues of low service availability (for example, data and storage availability) and imbalanced workload (that lead to low storage utilization) are strongly correlated with the number of social contacts. The studies [33,34] show that the small friend set is a cause of low service availability as well as poor storage utilization. However, it is worth noting that these findings are drawn in the context of exogenous social contacts. We show that, for the given utility function in symmetric social storage, if agents are allowed to select their storage partners, then each agent wants to form a social connection with only one other agent, or in other words, each agent has only one neighbor. We infer that if agents select their partners by looking at their cost-benefit trade-offs, then the issues discussed above are more significant than in the context of exogenous social contacts.

We believe that the analysis of storage availability and network formation performed by us have several advantages from the point of view of storage providers (for example, BuddyBackup, CrashPlan, Friendstore). The analysis of network stability may help design efficient strategies related to data redundancy that suggest how many data pieces are needed on the storage space provided by partners in order to achieve the required level of data availability. It also helps design efficient workload strategies to maximize storage utilization. One of the advantages of endogenous network formation is that it provides more control to agents on their data and in selecting their storage partners. Further, our approach of analysis of network stability is useful for the agents who are part of the Friendstore storage system—it is easy for them to calculate their maintainable capacity [2] so as to maximize their storage utilization and data reliability.

### 5.2. Limitations

Despite the above advantages and implications, our study has several limitations. Firstly, our social storage model stands on the assumption (similar to various network formation models [35–38]) that agents have complete information about the network structure. Though we do not require this assumption during network formation as, owing to Theorem 5, links form (at most) pairwise, this assumption is crucial for our analysis in Section 3 on closeness and distances. Secondly, although the proposed utility model captures various parameters essential for understanding social storage cloud formation, we cannot rule out that parameters like online availability of agents, trust between them, and the bandwidth they have may influence the network formation.

### 5.3. Future Scope

In this paper, though the utility function we propose is for heterogeneous agents, our analysis is limited to homogeneous agents (or symmetric social storage cloud systems). In the case of heterogeneous agents, it would be interesting to see how externalities will influence social storage cloud formation. Our analysis of storage availability will also be more relevant in this setting.

Further, we can enrich the utility model proposed in this paper by taking the above mentioned parameters (that is, online availability, bandwidth, and trust) into account. One can then study social storage cloud formation with both complete and incomplete information. For example, in the incomplete information setting, agents know neither the network structure nor the online availability and bandwidth of others. Analysis in this context will give more insight into social storage cloud formation.

**Author Contributions:** All authors have contributed equally to all aspects of this paper.

**Funding:** This research received no external funding.

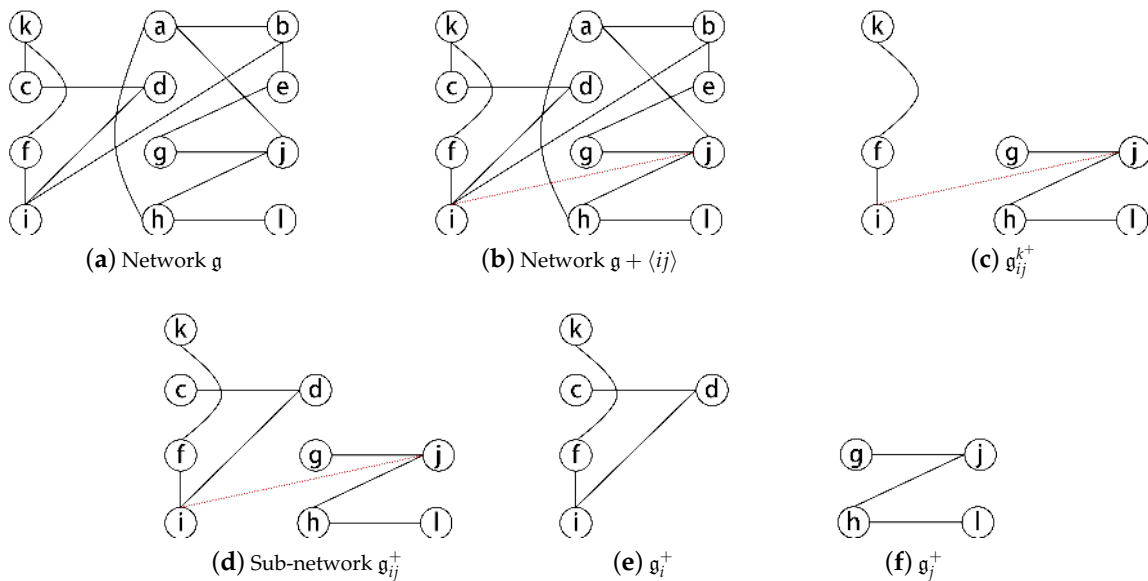
**Conflicts of Interest:** The authors declare no conflict of interest.

**Acknowledgments:** The authors would like to acknowledge the editor for the help and the anonymous reviewers for their valuable comments and suggestions, which have improved the quality of this paper.

**Appendix A. Example**

**Example A1.** Consider networks  $\mathbf{g}$  and  $\mathbf{g} + \langle ij \rangle$ , as shown in Figure A1a,b, respectively. The newly added link  $\langle ij \rangle$  in  $\mathbf{g}$  increases the closeness of agents  $k, c, d, f, g, j, i, h$ , and  $l$  in  $\mathbf{g} + \langle ij \rangle$ . For instance, from Equation (1), we have  $\Phi_k(\mathbf{g}) = 4.60$  and  $\Phi_k(\mathbf{g} + \langle ij \rangle) = 4.87$ . However, there is no change in the closeness of agents  $a, b$ , and  $e$ . For instance,  $\Phi_a(\mathbf{g}) = 6.17 = \Phi_a(\mathbf{g} + \langle ij \rangle)$ . Figure A1c shows the shortest-path-network  $\mathbf{g}_{ij}^{k+}$  for agent  $k$ , where the set  $L_k^+$  consists of agents  $j, g, h$ , and  $l$ . This suggests that the newly added link  $\langle ij \rangle$  in  $\mathbf{g}$  brings agents  $j, g, h$ , and  $l$  close to  $k$  in  $\mathbf{g} + \langle ij \rangle$ , and therefore,  $\Phi_k(\mathbf{g}) < \Phi_k(\mathbf{g} + \langle ij \rangle)$ . Figure A1d represents the shortest-path-network  $\mathbf{g}_{ij}^+$  that satisfies Definition 4. In this case,  $\mathbf{g}_{ij}^+$  is the union of  $\mathbf{g}_{ij}^{k+}, \mathbf{g}_{ij}^{c+}, \mathbf{g}_{ij}^{d+}, \mathbf{g}_{ij}^{f+}, \mathbf{g}_{ij}^{g+}, \mathbf{g}_{ij}^{h+}, \mathbf{g}_{ij}^{i+}, \mathbf{g}_{ij}^{j+}$ , and  $\mathbf{g}_{ij}^{l+}$ .

Figures A1e,f show the induced subnetworks  $\mathbf{g}_i^+$  and  $\mathbf{g}_j^+$  of  $\mathbf{g}_{ij}^+$ , respectively. The induced subnetworks  $\mathbf{g}_i^+$  and  $\mathbf{g}_j^+$  are as per Definition 5. In  $\mathbf{g}_i^+$ , the distances between agent  $i$ , and other agents  $k, c, d$ , and  $f$  are the same in networks  $\mathbf{g}$  and  $\mathbf{g} + \langle ij \rangle$ . We have  $d_{ik}(\mathbf{g}) = d_{ik}(\mathbf{g} + \langle ij \rangle) = 2$ ,  $d_{ic}(\mathbf{g}) = d_{ic}(\mathbf{g} + \langle ij \rangle) = 2$ ,  $d_{id}(\mathbf{g}) = d_{id}(\mathbf{g} + \langle ij \rangle) = 1$ , and  $d_{if}(\mathbf{g}) = d_{if}(\mathbf{g} + \langle ij \rangle) = 1$ . Similarly, in  $\mathbf{g}_j^+$ , the distances between agent  $j$ , and other agents  $g, h$ , and  $l$  are the same in network  $\mathbf{g}$  and  $\mathbf{g} + \langle ij \rangle$ ;  $d_{jg}(\mathbf{g}) = d_{jg}(\mathbf{g} + \langle ij \rangle) = 1$ ,  $d_{jh}(\mathbf{g}) = d_{jh}(\mathbf{g} + \langle ij \rangle) = 1$ , and  $d_{jl}(\mathbf{g}) = d_{jl}(\mathbf{g} + \langle ij \rangle) = 2$ .



**Figure A1.** Induced subnetworks of  $\mathbf{g} + \langle ij \rangle$ .

**Appendix B. Results Owing to Link Deletion**

*Appendix B.1. Effect of Closeness on Distances of Agents not Involved in Link Deletion*

Given  $k$  such that  $\Phi_k(\mathbf{g}) > \Phi_k(\mathbf{g} - \langle ij \rangle)$ , we use  $L_k^-$  to denote the set of all  $l \in \mathbf{g}$  such that all shortest paths from  $k$  to  $l$  in  $\mathbf{g} - \langle ij \rangle$  contain  $\langle ij \rangle$ . We use  $l_k^-$  to denote an agent in  $L_k^-$ .

**Proposition A1.** Let  $i \neq j \neq k, l \neq k, i \neq l$ , and  $\Phi_k(\mathbf{g} - \langle ij \rangle) < \Phi_k(\mathbf{g})$ . If  $d_{ik}(\mathbf{g}) < d_{ij}(\mathbf{g}) \leq d_{kl}(\mathbf{g})$ , then  $d_{ik}(\mathbf{g} - \langle ij \rangle) = d_{ik}(\mathbf{g})$ .

We skip the proof as it is similar to the proof of Proposition 1.

**Definition A1.** Suppose  $\langle ij \rangle \in \mathfrak{g}$  and  $k$  is an agent such that  $\Phi_k(\mathfrak{g}) > \Phi_k(\mathfrak{g} - \langle ij \rangle)$ . A  $(k, -ij)$ -shortest-path-network containing  $\langle ij \rangle$ ,  $\mathfrak{g}_{ij}^{k-}$ , is a subnetwork of  $\mathfrak{g}$  that consists of all shortest paths from  $k$  to  $l_k^-$  in  $\mathfrak{g}$ , which contain  $\langle ij \rangle$ , for all  $l_k^- \in L_k^-$ .

**Definition A2.** An (all  $k, -ij$ )-shortest-path-network,  $\mathfrak{g}_{ij}^-$  is  $\bigcup_{\substack{k \in \mathfrak{g}, \\ \Phi_k(\mathfrak{g}) > \Phi_k(\mathfrak{g} - \langle ij \rangle)}} \mathfrak{g}_{ij}^{k-}$ , the smallest network that contains all  $(k, -ij)$ -shortest-path-networks.

**Definition A3.** A sub- $(i, -)$ -network,  $\mathfrak{g}_i^-$  of  $\mathfrak{g}_{ij}^-$ , is the induced subnetwork of  $\mathfrak{g}_{ij}^-$  consisting of all agents  $k \in \mathfrak{g}_{ij}^-$  such that  $d_{ik}(\mathfrak{g}) = d_{ik}(\mathfrak{g} - \langle ij \rangle)$ . Similarly, we define the sub- $(j, -)$ -network,  $\mathfrak{g}_j^-$  of  $\mathfrak{g}_{ij}^-$ , as the induced subnetwork of  $\mathfrak{g}_{ij}^-$  consisting of all agents  $l \in \mathfrak{g}_{ij}^-$  such that  $d_{jl}(\mathfrak{g}) = d_{jl}(\mathfrak{g} - \langle ij \rangle)$ .

**Proposition A2.** Let  $k, \hat{k} \in \mathfrak{g}_i^-$ . Then, for all  $k, \hat{k} \in \mathfrak{g}_i^-$ ,  $d_{k\hat{k}}(\mathfrak{g}) = d_{k\hat{k}}(\mathfrak{g} - \langle ij \rangle)$ .

We skip the proof as it is similar to the proof of Proposition 2.

### Appendix B.2. Effect of Link Deletion on Storage Availability

We discuss the effect of link deletion on agents' storage availability.

**Lemma A1.** Suppose  $\langle ij \rangle \in \mathfrak{g}$  and  $k \notin \mathfrak{g}_{ij}^-$ . Then,  $\alpha_{ik}(\mathfrak{g}) = \alpha_{ik}(\mathfrak{g} - \langle ij \rangle)$  and  $\alpha_{jk}(\mathfrak{g}) = \alpha_{jk}(\mathfrak{g} - \langle ij \rangle)$ .

**Proof.** The proof follows from Proposition A1 and is in similar lines to the proof of Lemma 5.  $\square$

**Lemma A2.** Suppose agents  $i, j, k$ , and  $l$  are such that  $i \neq j, j \neq k, i \neq l$ , and  $k \neq l$ . (Agents  $i$  and  $k$  may be the same, and agents  $j$  and  $l$  may be the same). Suppose  $\langle ij \rangle \in \mathfrak{g}, k \in \mathfrak{g}_i^-,$  and  $l \in \mathfrak{g}_j^-$ . Then,

1.  $\alpha_{kl}(\mathfrak{g}) > \alpha_{kl}(\mathfrak{g} - \langle ij \rangle)$ .
2. If  $i \neq k$ , then  $\alpha_{ik}(\mathfrak{g}) < \alpha_{ik}(\mathfrak{g} - \langle ij \rangle)$ . Similarly, if  $j \neq l$ , then  $\alpha_{jl}(\mathfrak{g}) < \alpha_{jl}(\mathfrak{g} - \langle ij \rangle)$ .

**Proof.** The proof of 1 is in similar lines to the proof of 1 of Lemma 6. The proof of 2 follows from Proposition A2.  $\square$

**Lemma A3.** Let  $k$  and  $\hat{k}$  be agents in  $\mathfrak{g}_i^-$ . Then,  $\alpha_{k\hat{k}}(\mathfrak{g}) = \alpha_{k\hat{k}}(\mathfrak{g} - \langle ij \rangle)$  and  $\alpha_{\hat{k}k}(\mathfrak{g}) = \alpha_{\hat{k}k}(\mathfrak{g} - \langle ij \rangle)$ .

**Proof.** The proof follows from Lemma A2.  $\square$

**Theorem A1.** Suppose agents  $i, j, k$ , and  $l$  are such that  $i \neq j$  and  $k \neq l$ . Suppose  $\langle ij \rangle \in \mathfrak{g}$ .

Then,  $\gamma_i(\mathfrak{g}) < \gamma_i(\mathfrak{g} - \langle ij \rangle)$  if and only if  $\frac{\prod_{l \in \mathfrak{g}_j^-} (1 - \alpha_{ik}(\mathfrak{g}))}{\prod_{l \in \mathfrak{g}_j^-} (1 - \alpha_{il}(\mathfrak{g} - \langle ij \rangle))} < \frac{\prod_{k \in \mathfrak{g}_i^-} (1 - \alpha_{ik}(\mathfrak{g} - \langle ij \rangle))}{\prod_{k \in \mathfrak{g}_i^-} (1 - \alpha_{il}(\mathfrak{g}))}$ .

In addition,  $\gamma_i(\mathfrak{g}) < \gamma_i(\mathfrak{g} - \langle ij \rangle)$  if and only if  $\frac{\prod_{l \in \mathfrak{g}_j^-} (\alpha_{ik}(\mathfrak{g}))}{\prod_{l \in \mathfrak{g}_j^-} (\alpha_{il}(\mathfrak{g} - \langle ij \rangle))} > \frac{\prod_{k \in \mathfrak{g}_i^-} (\alpha_{ik}(\mathfrak{g} - \langle ij \rangle))}{\prod_{k \in \mathfrak{g}_i^-} (\alpha_{il}(\mathfrak{g}))}$ .

**Proof.** The proof follows from Lemmas A1–A3.  $\square$

### Appendix C. Proof of Lemma 6

Proof of 2 follows from Proposition 1.

Proof of 1 is as follows.



As  $k \in \mathfrak{g}_i^+$  and  $l \in \mathfrak{g}_j^+$ ,  $d_{kl}(\mathfrak{g}) > d_{kl}(\mathfrak{g} + \langle ij \rangle)$ , hence,  $\Phi_k(\mathfrak{g}) < \Phi_k(\mathfrak{g} + \langle ij \rangle)$  and  $\Phi_l(\mathfrak{g}) < \Phi_l(\mathfrak{g} + \langle ij \rangle)$ .

It is easy to see that network  $\mathfrak{g}$  in Figure A2a is the one where adding link  $\langle ij \rangle$  leads to the maximum increment in  $l$ 's closeness and the minimum decrements in the distances between  $l$  and  $k_m$ , ( $m = 1, 2, \dots, n - 4$ , where  $n$  is the number of agents in  $\mathfrak{g}$ ), the maximum and the minimum being across all network structures.

We consider two cases,  $j \neq l$  and  $j = l$ .

Suppose  $j \neq l$ . Consider the network  $\mathfrak{g}$ , as shown in Figure A2a.

$$\begin{aligned} \text{From Equation (1), } \Phi_l(\mathfrak{g}) &= \frac{1}{d_{ij}(\mathfrak{g})} + \frac{1}{d_{lx}(\mathfrak{g})} + \frac{1}{d_{li}(\mathfrak{g})} + \sum_{\substack{k \in \mathfrak{g}, \\ d_{kl}(\mathfrak{g})=4}} \frac{1}{d_{kl}(\mathfrak{g})} \\ &= 1 + \frac{1}{2} + \frac{1}{3} + \frac{n-4}{4} = \frac{3n+10}{12}. \end{aligned}$$

Without loss of generality, let  $k = k_m$  for some  $m \in \{1, 2, \dots, n - 4\}$ . Then, from Equation (2),  $\alpha_{kl}(\mathfrak{g}) = \frac{(\frac{1}{4})}{\Phi_l(\mathfrak{g})} = \frac{3}{3n+10}$ .

If agents  $i$  and  $j$  add a direct link in  $\mathfrak{g}$ , we have network  $\mathfrak{g} + \langle ij \rangle$ , as shown in Figure A2b. Then, from Equations (1) and (2), we have  $\Phi_l(\mathfrak{g} + \langle ij \rangle) = \frac{n+2}{3}$  and  $\alpha_{kl}(\mathfrak{g} + \langle ij \rangle) = \frac{1}{n+2}$ .

From the above, clearly,  $\alpha_{kl}(\mathfrak{g}) < \alpha_{kl}(\mathfrak{g} + \langle ij \rangle)$ .

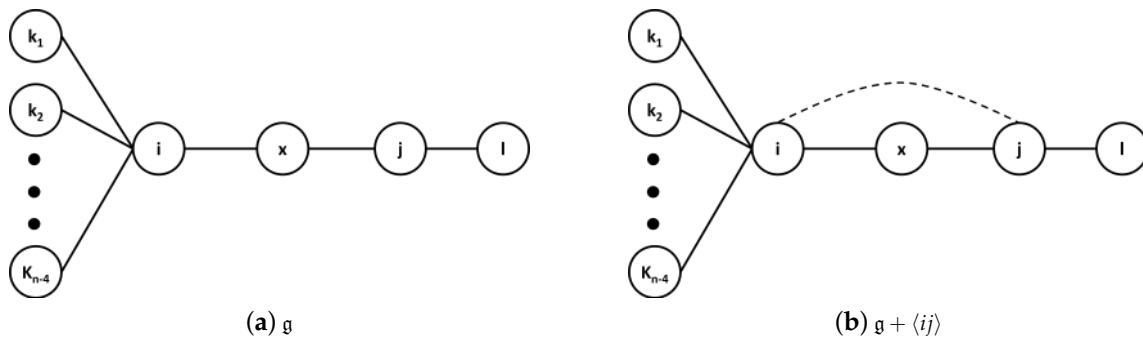


Figure A2. Network structure  $\mathfrak{g}$  and  $\mathfrak{g} + \langle ij \rangle$  with  $n$  agents.

Now, suppose  $l = j$ .

From Equations (1) and (2) applied to Figure A2a,b, we have  $\Phi_j(\mathfrak{g}) = \frac{2n+7}{6}$ ,  $\alpha_{kj}(\mathfrak{g}) = \frac{2}{2n+7}$ ,  $\Phi_j(\mathfrak{g} + \langle ij \rangle) = \frac{n+2}{2}$ , and  $\alpha_{kj}(\mathfrak{g} + \langle ij \rangle) = \frac{1}{n+2}$ .

It is easy to see that  $\alpha_{kj}(\mathfrak{g}) < \alpha_{kj}(\mathfrak{g} + \langle ij \rangle)$  in this case as well. This completes the proof of 1.  $\square$

### Appendix D. Experimental Results

We conduct random experiments to answer the following question. Though agents form links and backup their data with adjacent agents, can any agent still lose its data? From Theorem 5, the null network is the unique pairwise stable network when the cost to add links is "high", that is,  $\zeta \geq \frac{\beta}{\lambda}$ , and pairs of agents with links between them is the unique stable network otherwise. Therefore, as far as formation of networks is concerned, we always obtain one of these two networks, depending on the values of  $\zeta$ ,  $\beta$ , and  $\lambda$ . In our experiment, we randomly generate networks of the second type, namely, pairwise links. We generate such networks on 30 agents and consider 150 random scenarios, by generating 10 random networks, 5 different sets of randomly chosen agents whose storage disks fail, for 3 cases,  $\lambda = 1\%$ ,  $2\%$ , and  $4\%$ . Our assumption on the values of  $\lambda$  is based on data from Backblaze<sup>9</sup> on hard drive failure rates. Interestingly, in none of the random cases we generated did agents on the two sides of a link fail at the same time.

<sup>9</sup> <https://www.backblaze.com/blog/backblaze-hard-drive-stats-q1-2019/> (accessed on 04 September 2019).

## References

1. Chard, K.; Bubendorfer, K.; Caton, S.; Rana, O.F. Social cloud computing: A vision for socially motivated resource sharing. *IEEE Trans. Serv. Comput.* **2012**, *5*, 551–563. [[CrossRef](#)]
2. Tran, N.; Chiang, F.; Li, J. Efficient cooperative backup with decentralized trust management. *Trans. Storage* **2012**, *8*, 8:1–8:25. [[CrossRef](#)]
3. Gracia-Tinedo, R.; Sánchez-Artigas, M.; García-López, P. F2Box: Cloudifying F2F storage systems with high availability correlation. In Proceedings of the 2012 IEEE Fifth International Conference on Cloud Computing (CLOUD), Honolulu, HI, USA, 24–29 June 2012; pp. 123–130. [[CrossRef](#)]
4. Moreno-Martínez, A.; Gracia-Tinedo, R.; Sánchez-Artigas, M.; Garcia-Lopez, P. FRIENDBOX: A cloudified F2F storage application. In Proceedings of the 2012 IEEE 12th International Conference on Peer-to-Peer Computing (P2P), Tarragona, Spain, 3–5 September 2012; pp. 75–76. [[CrossRef](#)]
5. Nguyen, T.D.; Li, J. BlockParty: Cooperative offsite backup among friends. In Proceedings of the 4th USENIX Symposium on Networked Systems Design & Implementation, Cambridge, MA, USA, 11–13 April 2007.
6. Tran, N.; Li, J.; Subramanian, L.; Chow, S.S. Optimal Sybil-resilient node admission control. In Proceedings of the 2011 Proceedings IEEE INFOCOM, Shanghai, China, 10–15 April 2011; pp. 3218–3226. [[CrossRef](#)]
7. Zuo, X.; Iamnitchi, A. A survey of socially aware peer-to-peer systems. *ACM Comput. Surv.* **2016**, *49*, 9:1–9:28. [[CrossRef](#)]
8. Freeman, L.C. Centrality in social networks conceptual clarification. *Soc. Netw.* **1978**, *1*, 215–239. [[CrossRef](#)]
9. Borgatti, S.P.; Everett, M.G. A Graph-theoretic perspective on centrality. *Soc. Netw.* **2006**, *28*, 466–484. [[CrossRef](#)]
10. Boldi, P.; Vigna, S. Axioms for centrality. *Internet Math.* **2014**, *10*, 222–262. [[CrossRef](#)]
11. Bloch, F.; Jackson, M.O.; Tebaldi, P. Centrality measures in networks. *arXiv* **2017**, arXiv:1608.05845.
12. Skibski, O.; Sosnowska, J. Axioms for distance-based centralities. In Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018; pp. 1218–1225.
13. Rao, N.S.V.; Ma, C.Y.T.; He, F.; Yau, D.K.Y.; Zhuang, J. Cyber-physical correlation effects in defense games for large discrete infrastructures. *Games* **2018**, *9*, 52. [[CrossRef](#)]
14. Altman, E.; Kameda, H.; Hosokawa, Y. Nash equilibria in load balancing in distributed computer systems. *Int. Game Theory Rev.* **2002**, *04*, 91–100. [[CrossRef](#)]
15. Hausken, K. Information sharing among cyber hackers in successive attacks. *Int. Game Theory Rev.* **2017**, *19*, 1750010-1–1750010-33. [[CrossRef](#)]
16. Timmer, J.; Scheinhardt, W. Customer and cost sharing in a Jackson network. *Int. Game Theory Rev.* **2018**, *20*, 1850002-1–1850002-10. [[CrossRef](#)]
17. Pilling, R.; Chang, S.C.; Luh, P.B. Shapley value-based payment calculation for energy exchange between micro- and utility grids. *Games* **2017**, *8*, 45. [[CrossRef](#)]
18. Sanchez-Soriano, J. An overview on game theory applications to engineering. *Int. Game Theory Rev.* **2013**, *15*, 1340019-1–1340019-18. [[CrossRef](#)]
19. Dutta, B.; Jackson, M.O. On the formation of networks and groups. In *Networks and Groups: Models of Strategic Formation (Studies in Economic Design)*, 1st ed.; Dutta, B., Jackson, M.O., Eds.; Springer: Berlin/Heidelberg, Germany, 2003; Volume VIII, pp. 1–15.
20. Jackson, M.O. A survey of network formation models: Stability and efficiency. In *Group Formation in Economics: Networks, Clubs, and Coalitions*; Demange, G., Wooders, M., Eds.; Cambridge University Press: Cambridge, UK, 2005; pp. 11–57.
21. Marini, M.A. Games of coalition and network formation: A survey. In *Networks, Topology and Dynamics*, 1st ed.; Naimzada, A.K., Stefani, S., Torriero, A., Eds.; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 2009; Volume 613, pp. 67–93.
22. Borkotokey, S.; Gogoi, L.; Sarangi, S. A survey of player-based and link-based allocation rules for network games. *Stud. Micro.* **2014**, *2*, 5–26. [[CrossRef](#)]
23. Mane, P.C.; Ahuja, K.; Krishnamurthy, N. Stability, efficiency, and contentedness of social storage networks. *Ann. Oper. Res.* **2019**, 1–32. [[CrossRef](#)]
24. Morrill, T. Network formation under negative degree-based externalities. *Internat. J. Game Theory* **2011**, *40*, 367–385. [[CrossRef](#)]

25. Möhlmeier, P.; Rusinowska, A.; Tanimura, E. A degree-distance-based connections model with negative and positive externalities. *J. Public Econ. Theory* **2016**, *18*, 168–192. [[CrossRef](#)]
26. Jackson, M.O.; Wolinsky, A. A strategic model of social and economic networks. *J. Econ. Theory* **1996**, *71*, 44–74. [[CrossRef](#)]
27. Opsahl, T.; Agneessens, F.; Skvoretz, J. Node centrality in weighted networks: Generalizing degree and shortest paths. *Soc. Networks* **2010**, *32*, 245–251. [[CrossRef](#)]
28. Marchiori, M.; Latora, V. Harmony in the small-world. *Phys. A* **2000**, *285*, 539–546. [[CrossRef](#)]
29. Lillibridge, M.; Elnikety, S.; Birrell, A.; Burrows, M.; Isard, M. A cooperative Internet backup scheme. In Proceedings of the Annual Conference on USENIX Annual Technical Conference, San Antonio, TX, USA, 9–14 June 2003; pp. 29–41.
30. Landers, M.; Zhang, H.; Tan, K.L. PeerStore: Better performance by relaxing in peer-to-peer backup. In Proceedings of the Fourth International Conference on Peer-to-Peer Computing, Zurich, Switzerland, 27–27 August 2004; pp. 72–79. [[CrossRef](#)]
31. Cox, L.P.; Murray, C.D.; Noble, B.D. Pastiche: Making backup cheap and easy. *SIGOPS Oper. Syst. Rev.* **2002**, *36*, 285–298. [[CrossRef](#)]
32. Jackson, M.O. *Social and Economic Networks*, 2nd ed.; Princeton University Press: Princeton, NJ, USA, 2010; pp. 215–216.
33. Sharma, R.; Datta, A.; DeH’Amico, M.; Michiardi, P. An empirical study of availability in friend-to-friend storage systems. In Proceedings of the 2011 IEEE International Conference on Peer-to-Peer Computing, Kyoto, Japan, 31 August–2 September 2011; pp. 348–351.
34. Zuo, X.; Blackburn, J.; Kourtellis, N.; Skvoretz, J.; Iamnitchi, A. The power of indirect ties in friend-to-friend storage systems. In Proceedings of the 14-th IEEE International Conference on Peer-to-Peer Computing, London, UK, 8–12 September 2014; pp. 1–5.
35. Bala, V.; Goyal, K. A noncooperative model of network formation. *Econometrica* **2000**, *68*, 1181–1229. [[CrossRef](#)]
36. Johnson, C.; Gilles, R. Spatial social networks. *Rev. Econ. Des.* **2000**, *5*, 273–299. [[CrossRef](#)]
37. Moscibroda, T.; Schmid, S.; Wattenhofer, R. Topological implications of selfish neighbor selection in unstructured peer-to-peer networks. *Algorithmica* **2011**, *61*, 419–446. [[CrossRef](#)]
38. Buechel, B. Network Formation with Closeness Incentives. In *Networks, Topology and Dynamics: Theory and Applications to Economics and Social Systems*, 1st ed.; Naimzada, A.K., Stefani, S., Torriero, A., Eds.; Lecture Notes in Economics and Mathematical Systems; Springer: Berlin/Heidelberg, Germany, 2009; Volume 613, pp. 95–109.



© 2019 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).