Technical Appendix for
"Media, Aggregators and the Link Economy:
Strategic Hyperlink Formation in Content Networks"

Chrysanthos Dellarocas
Zsolt Katona
William Rand

April 24, 2013

1 The anchor selection process

In Section 3 of the paper, we use a reduced form model to account for how consumers select their anchor node. In this section, we provide a formal micro-model of a process that provides justification for our assumptions. We also consider the possibility that some consumers may switch their anchor nodes when encountering a link to a higher quality site and show that including this feature in our model does not qualitatively change our results.

In common with other analyses of web-browsing behavior, we employ a Markovian model to abstract the anchor node selection process. Model states $i = 0, 1, .., N$ designate which site a consumer uses as her anchor node. State 0 corresponds to the situation where the consumer anchors herself at the outside option. We define transition probabilities from one state to another, representing the likelihood that consumers change their anchor node. We follow the behavior of one randomly selected consumer and measure the probabilities that the consumer anchors herself at a particular node. Let p_i define the probability that the consumer is anchored at site $i = 0, 1, ..N$; p_0 corresponds to the outside option. Let w_{ij} measure the transition probability from node i to j, that is, the probability that the consumer switches her anchor from i to j, given that her current anchor is j. We assume

$$\sum_{j=0}^{N} w_{ij} = 1$$

for any $0 \le i \le N$. Our goal is to solve for the probabilities p_i^* that keep the system in equilibrium. Formally, we require these probabilities to satisfy

$$\sum_{i=0}^{N} p_i^* = 1 \ and \ p_j^* = \sum_{i=0}^{N} p_i^* w_{ij} \tag{11}$$

for any $0 \le j \le N$, making $(p_0^*, p_1^*, ..., p_N^*)$ an eigenvector of the W matrix formed by the w_{ij}

transition probabilities.

First we model a setup without links and assume that consumers generally stick to their anchor nodes, but they give them up with some probability and move to the outside option. We assume that, once a consumer anchors herself at a site, the probability w_{i0} that she abandons it is inversely proportional to the content the consumer gets, that is, $w_{i0} = \nu / z_i$, for $1 \leq i \leq N$, where ν is a sufficiently small constant.[1] When consumers are in the outside option state (state 0), they enter the market again and choose among one of the sites $i = 1, .., N$ with equal probabilities. The probability that they choose any given site is inversely proportional to the attractiveness of the outside option, yielding $w_{0i} = \nu / \mu$, for $1 \leq i \leq N$. Observe that we assume that consumers do not directly switch from one anchor to another. Switching requires passing through the outside option, that is, $w_{ij} = 0$ for any $1 \leq i, j \leq N$ pair. The above process captures the notion that consumers experiment with different anchor nodes and stick to these sites depending on the content they obtain. If the content is not high, they are more likely to leave and go back to the outside option from where they start experimenting again, especially if the outside option is not very attractive. Alternatively, a transition to the outside option can be interpreted as consumer exit, whereas a transition from the outside option as a new consumer entry.[2]

The equilibrium condition (11) translates to

$$p_i^* = \left(1 - \frac{\nu}{z_i}\right) p_i^* + p_0^* \frac{\nu}{\mu}$$

for any $1 \leq i \leq N$, leading to $p_i^* / p_0^* = z_i / \mu$. Thus,

$$p_i^* = \frac{z_i}{\mu + z_1 + z_2 + ... + z_N}, \quad \text{and} \quad p_o^* = \frac{\mu}{\mu + z_1 + z_2 + ... + z_N}$$

which are the anchor traffic formulae we have assumed in Section 3 of the paper.

Now let us consider a case with links, in particular, the most representative case where all but one site are linking to one target, node 1. When a consumer encounters a link source, she will eventually access the link target's content even if she anchors at the link source. In the paper we assume that the presence of a link does not affect the choice of anchor node, because the consumer gets the same satisfaction even if she has to click to get to the content. Here, we relax this assumption and let consumers change their anchor node upon observing a link to another node. We assume that, when a consumer chooses her anchor node (transitioning from the outside option) and observes a link to another node, she changes her anchor node to the link target with probability α. With this additional assumption in place, the transition probabilites change from $w_{0i} = \frac{\nu}{\mu}$ to $w'_{0i} = (1 - \alpha)\frac{\nu}{\mu}$ for $2 \leq i \leq N$ and $w'_{01} = (1 + (N-1)\alpha)\frac{\nu}{\mu}$. The rest of the transition probabilites are unchanged. Using these new transition probabilities, we can calculate the equilibrium anchor selection probabilities, yielding

[1] In particular, we assume that $\nu < min(\mu/N, z_1, z_2, ..., z_N)$

[2] The assumption that consumers transition between anchor nodes only throught the outside option is not critical: the results would be very similar if consumers immediately moved to other anchor nodes when not satisfied with one. We decided to use this set of assumptions because it allows us to obtain the exact same functional form of anchor traffic we use in the paper.

$$p_1^* = \frac{z_1(1+(N-1)\alpha)}{\mu + z_1 + z_2 + ... + z_N}, \;\; p_i^* = \frac{z_i(1-\alpha)}{\mu + z_1 + z_2 + ... + z_N}, \;\; \text{and} \;\; p_o^* = \frac{\mu}{\mu + z_1 + z_2 + ... + z_N}$$

for any $2 \leq i \leq N$. For two nodes, where site 2 links to site 1 this results in

$$p_1^* = \frac{c_1(1+\alpha)}{\mu + 2c_1}, \;\; p_2^* = \frac{c_1(1-\alpha)}{\mu + 2c_1}.$$

The following result shows that, if we rewrite our basic model using the above probabilities, we get our original results with the simple substitution $\rho' = (1-\alpha)\rho$.

Proposition 9. *Using the functions $\underline{L}(\rho), \overline{NL}(\rho) \in [0,2]$ defined in Proposition 1 of the paper, we obtain that:*

1. *If $\mu \leq \overline{NL}(\rho(1-\alpha))$ then sites do not establish links in equilibrium and $c_i^* = c_j^* = c_{NL}$.*
2. *If $\mu \geq \underline{L}(\rho(1-\alpha))$ then there are two asymmetric equilibria where one site links to the other and*

$$c_T = \frac{1-\mu}{2} - \frac{\rho(1-\alpha)}{4} + \frac{\sqrt{(2-\rho(1-\alpha))(4\mu + 2 - \rho(1-\alpha))}}{4} >$$

$$c_S = \frac{(1-\alpha)\rho c_T}{2c_T + \mu}$$

3. *There is no equilibrium in pure strategies otherwise.*

Proof. From equation (3) of Section 3, substituting $T = 1, \; S = 2$ we obtain:

$$t_T^A = p_1^* = \frac{c_T(1+\alpha)}{\mu + 2c_T}, \;\; t_S^A = p_2^* = \frac{c_T(1-\alpha)}{\mu + 2c_T}.$$

The corresponding payoff functions of the link source and target nodes (see Section 4.1) then take the form:

$$\pi_S = \frac{c_T(1-\alpha)}{\mu + 2c_T}\rho c_S - \frac{k_S}{2}c_S^2 \qquad \pi_T = \frac{c_T(1+\alpha) + (1-\rho)c_T(1-\alpha)}{2c_T + \mu}c_T - \frac{k_T}{2}c_T^2 \tag{12}$$

which is equivalent to

$$\pi_S = \frac{c_T}{\mu + 2c_T}\rho(1-\alpha)c_S - \frac{k_S}{2}c_S^2 \qquad \pi_T = \frac{c_T + (1-\rho(1-\alpha))c_T}{2c_T + \mu}c_T - \frac{k_T}{2}c_T^2 \tag{13}$$

Comparing (13) with equation (5) of Section 4.1 we see that the two are equivalent with the simple substitution $\rho' = (1-\alpha)\rho$. Therefore, all results that rely on equation (5) also apply to this setting. □

The above result shows that a higher probability α of changing one's anchor node to the link target is mathematically equivalent to a lower ρ, that is, it reduces the source node's ability to monetize the link target's content. From this result it follows that all results of this paper remain qualitatively unchanged if we assume that users can change anchor nodes to the link target of their current anchor node. With reference to Figure 2, for a given ρ, linking becomes less likely as α increases, because then the effective $\rho' = \rho(1-\alpha)$ declines. For $\alpha = 1$ ($\rho' = 0$), we either have a unique no-link equilibrium (low μ) or both the linking and non-linking equlibria coexist (high μ).

2 Endogenous ρ

Throughout the paper, we assume that the proportion of anchor traffic a link source is able to keep is fixed and does not depend on the relative content levels. Here, we relax this assumption and allow ρ to depend on the link target's content. In this way, we can capture the possibility that visitors are more likely to click on a link when the link target's content is higher. We analyze the case with two nodes. To ensure analytical tractability, we assume a simple linear dependence. In particular, when site 1 links to site 2 we assume that the probability that a user arriving to site 1 does not click is

$$\rho(\tau) = \max(\rho_0(1 - \tau c_T), 0),$$

where τ measures how strongly the probability of clicking is affected by the link target's content. Naturally, we assume that a higher content attracts more clicks, which is why the probability that consumers stay at the link source is a decreasing function of the link target's content. Note that the linear formulation would result in negative probabilities if τ or c_T are high, thus we assume that consumers always click in these cases. However, we will only derive results for parameter values with an interior solution, therefore we assume $\tau < 1/2$. This formulation allows direct comparison with our original setup, since $\tau = 0$ is equivalent to our basic model with a fixed $\rho = \rho_0$.

With the exception of the probability of clicking, our setup is equivalent to our basic model, hence the content levels are the same when sites do not link to each other:

$$c_1^* = c_2^* = c_{NL} = \frac{3 - 4\mu + \sqrt{9 + 8\mu}}{8}$$

Similarly to our main results, we can show that a linking equilibrium is possible.

Proposition 10. *There exist thresholds* $\underline{L}(\rho_0, \tau), \overline{NL}(\rho_0, \tau) \in [0, 2]$ *such that:*

1. *If* $\mu \leq \overline{NL}(\rho_0, \tau)$ *then sites do not establish links in equilibrium and* $c_i^* = c_j^* = c_{NL}$.

2. *If* $\mu \geq \underline{L}(\rho_0, \tau) > 0$ *then there are two asymmetric equilibria where one site links to the other and*

$$c_T = \frac{4 - 2\rho_0 + (3\rho_0\tau - 4)\mu + \sqrt{(4 - 2\rho_0 - \rho_0\tau\mu)(4 - 2\rho_0 + (8 - 9\rho_0\tau)\mu)}}{8(1-\tau)} >$$

$$c_S = \frac{\rho_0(1 - \tau c_T)c_T}{2c_T + \mu}$$

3. There is no equilibrium in pure strategies otherwise.

Proof. As in the basic model, we examine when the linking and no-linking equilibria are feasible. Since we already calculated the possible content levels in the no-linking case, let us consider the linking equilibrium. When site i links to site j, site j's payoff is

$$\pi_{j\leftarrow i} = \frac{2+\rho_0(\tau c_j - 1)}{2c_j+\mu}c_j^2 - \frac{1}{2}c_j^2.$$

Differentiating $\pi_{j\leftarrow i}$ with respect to c_j yields that site j will invest c_T in content if site i links to it (as given in the proposition). Then, differentiating

$$\pi_{i\rightarrow j} = \frac{c_j\rho_0(1-\tau c_j)}{c_j+c_j+\mu}c_i - \frac{1}{2}c_i^2$$

with respect to c_i yields that site i will invest

$$b_{i\rightarrow j}(c_j) = \frac{\rho_0(1-\tau c_j)c_j}{2c_j+\mu}$$

in content if it links to j, yielding the stated $c_i = c_S$ if we plug $c_j = c_T$.

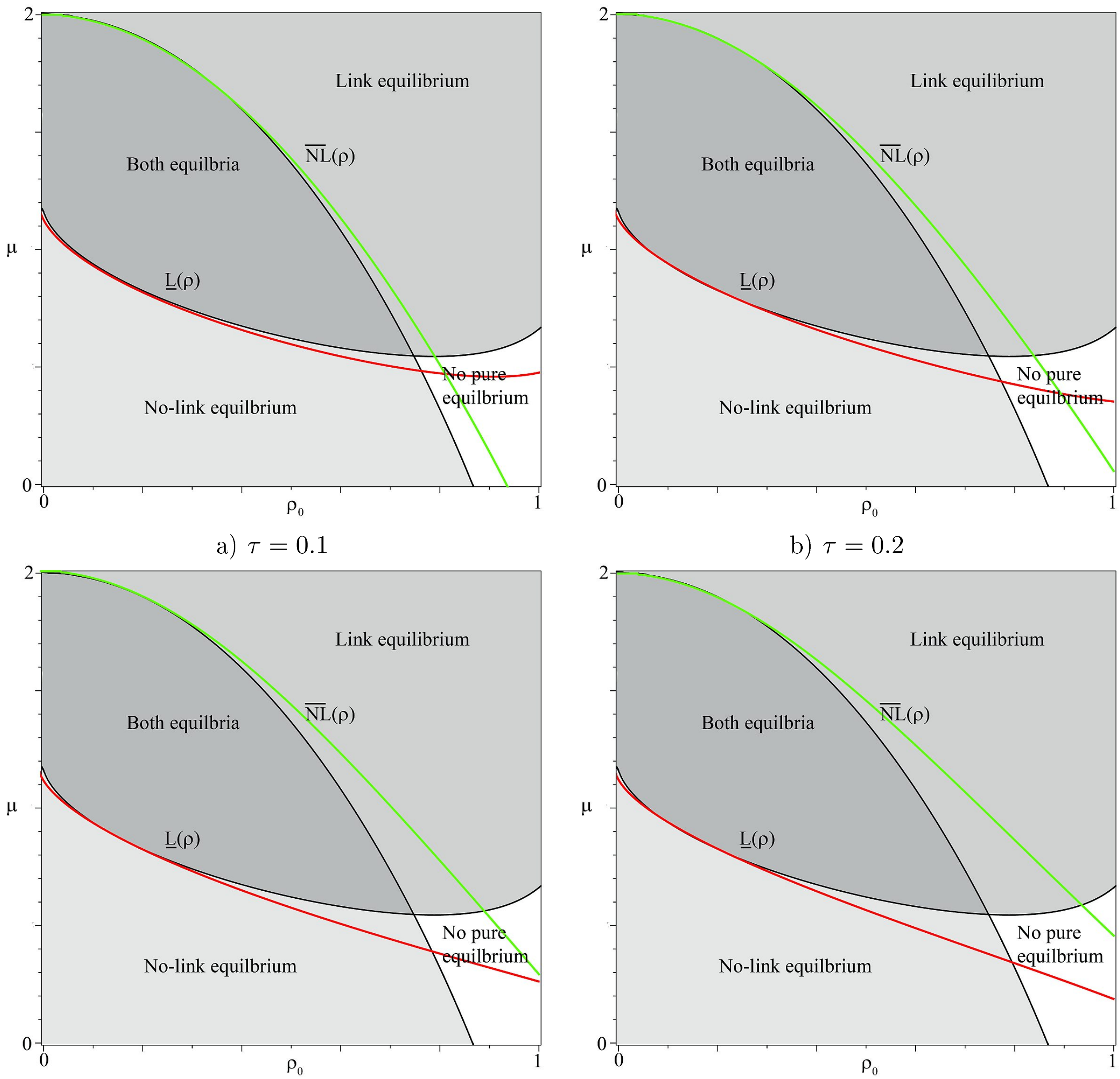
As in the basic case, to check whether sites have any incentive to deviate from the potential equilibria, we examine whether the no linking best response would yield higher profits in the linking case and whether the linking best response would yield higher profits in the no-link case. In the first case, the linking equilibrium holds iff

$$\pi_i(b_i(c_T), c_T) \leq \pi_S := \pi_{i\rightarrow j}(c_S, c_T).$$

Let $\underline{L}(\rho_0, \tau)$ denote the value of μ where the above holds with equality. The above inequality holds for high values of μ, yielding that the linking equilibria exists iff $\mu \geq \underline{L}(\delta, \rho)$. For $\mu = 0$, one can check that Similarly, let $\overline{NL}(\rho_0, \tau)$ denote the value of μ for which $\pi_i(c_{NL}, c_{NL}) = \pi_{i\rightarrow j}(b_{i\rightarrow j}(c_{NL})_S, c_{NL})$. Sites do not have an incentive to deviate from the no-link equilibrium iff $\mu \leq \overline{NL}(\delta, \rho)$, completing the proof. □

The results are not substantially different from our basic model. The equilibrium regions are similar to the case when ρ does not depend on the content levels. Furthermore, the linking equilibrium is not feasible when $\mu = 0$. Figure 2 shows how the equilibrium regions change as τ increases. As ρ depends more on the target's content, initially both the linking and non-linking equilibrium regions expand.

Both types of equilibria are more sustainable because of the following intuition. The clicking probability's dependence on the target's content creates more incentive for the target to invest in content. Therefore, in the linking equilibria it is less likely that the source finds it profitable to deviate by giving up the link and directly competing with the target. However, the no-linking equilibrium is also more sustainable, because both sites have lower incentives to deviate from a symmetric no-



a) $\tau = 0.1$ b) $\tau = 0.2$ c) $\tau = 0.3$ d) $\tau = 0.4$

Figure 8: Equilibrium regions as a function of ρ_0 and τ. The gray regions depict the parameter regions where the different equilibria are feasible in the basic model, that is, when $\tau = 0$. The red and green plots show how $\underline{L}(\rho_0, \tau)$ and $,\overline{NL}(\rho_0, \tau)$ change respectively as τ increases.

linking equilibrium by linking to the other one since lowering the content has a more pronounced negative effect by driving away visitors.

Corollary 2. *Both the target's content and profit are increasing in τ. The source's profit and content are decreasing in τ when ρ_0 is small.*

Not surprisingly, if the target's content has a positive effect on the click-through probability of the source's visitors the target is better off. Also, due to the increased incentives, it will invest more in its content. The source's content and profit are decreasing if ρ_0 is low, that is, when not many visitors stay at the source's site. However, when ρ_0 is sufficiently high this can change. The target's increased content level makes the source invest more in its own content to take advantage of the higher amount of traffic attracted. Since a high proportion of this traffic stays at the source this effect outweighs the direct effect of losing visitors to the higher content target.

3 Sequential moves

Throughout the paper we assume that sites make their decisions simultaneously. In this Section we analyze a sequential-move version of the basic two-player model. Site 1, the leader, first decides its content level and whether to link to site 2, the follower.[3] The follower then decides its content level and whether to link to the leader. Let us first analyze the case without links, where sites can only make content investments, but cannot establish links. We use backward induction and consider the follower's decision. As in the basic model site 2's payoff function is given by

$$\pi_2 = \frac{c_2}{c_1 + c_2 + \mu} c_2 - \frac{c_2^2}{2}$$

where we set $\delta = k_1 = k_2 = 1$. Differentiating π_2 with respect to c_2 and setting it to 0 yields that the best response for site 2:

$$b_2(c_1) = \frac{1 - 2\mu - 2c_1 + \sqrt{1 + 4\mu + 4c_1}}{2}.$$

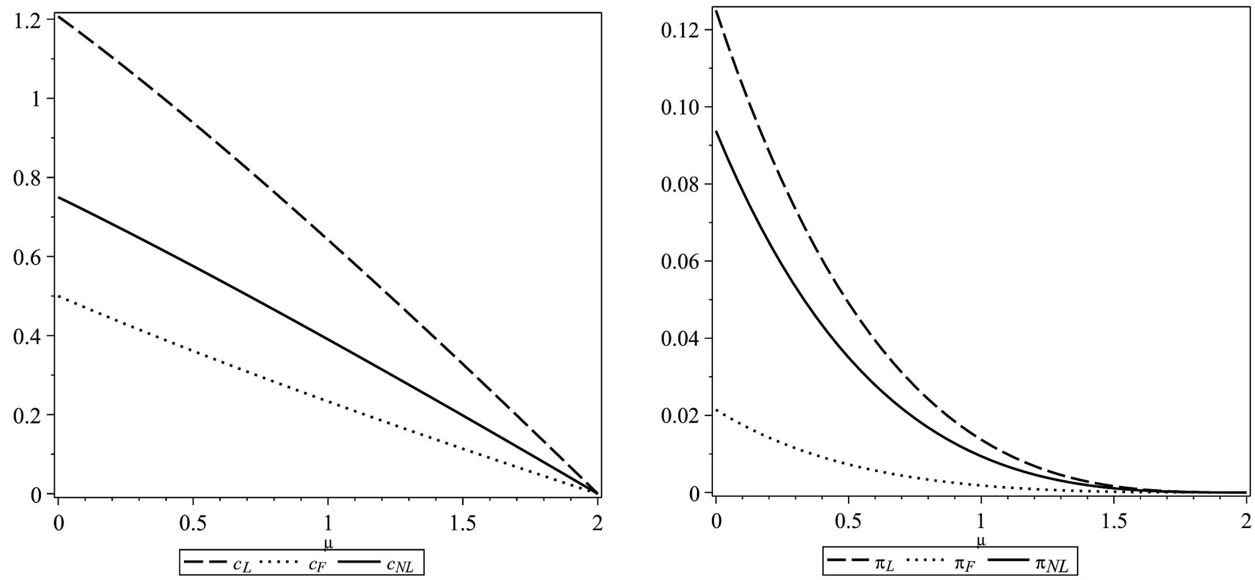
Plugging this into the leader's payoff function we get

$$\pi_1 = \frac{c_1^2}{2} \cdot \frac{3 - \sqrt{1 + 4\mu + 4c_1}}{1 + \sqrt{1 + 4\mu + 4c_1}}.$$

This function has a unique maximum yielding

$$c_1^* = c_L = \frac{\left(\frac{A}{3} + \frac{10}{3A} + \frac{1}{3}\right)^2 - 4\mu - 1}{4},$$

[3]Since we assume sequential moves, the follower has not yet produced its content when the leader makes its move. The way to interpret the leader's linking strategy is that site 1 *commits* to linking to the follower once the latter produces content (whose quality site 1 can perfectly predict).



(a) Content (b) Payoffs

Figure 9: Equilibrium content levels and payoffs when linking is not allowed and players move sequentially

where

$$A = \left(28 + 54\mu + 6\sqrt{81\mu^2 + 84\mu - 6}\right)^{1/3}.$$

which, in turn implies

$$c_2^* = c_F = \frac{1 - 2\mu - 2c_L + \sqrt{1 + 4\mu + 4c_L}}{2}.$$

Note that we naturally get

$$c_L > c_{NL} > c_F \text{ and } \pi_L > \pi_{NL} > \pi_F$$

where c_{NL} is the content level produced by two symmetric sites that cannot link to each other. That is, the leader has an advantage and can sustain a higher content level, thus making higher profits by attracting more traffic. Figure 9 depicts the equilibrium content levels and payoffs as functions of μ.

Moving on to the case with the possibility of links, we can expect that the ability of the leader to dominate the market without links, makes it more attractive for the follower to deviate from a non-linking equilibrium and link to the leader. The complete analytical characterization of equilibria is not possible due to the complexity of the expressions involved, therefore we derive the following results using numerical analysis.

Proposition 11. *There exist thresholds* $\underline{LT}(\mu), \overline{LT}(\mu) \in [0, 1]$ *such that:*

1. *If* $\rho < \underline{LT}(\mu)$ *then sites do not establish links in equilibrium and*

$$c_1^* = c_L \text{ and } c_2^* = c_F.$$

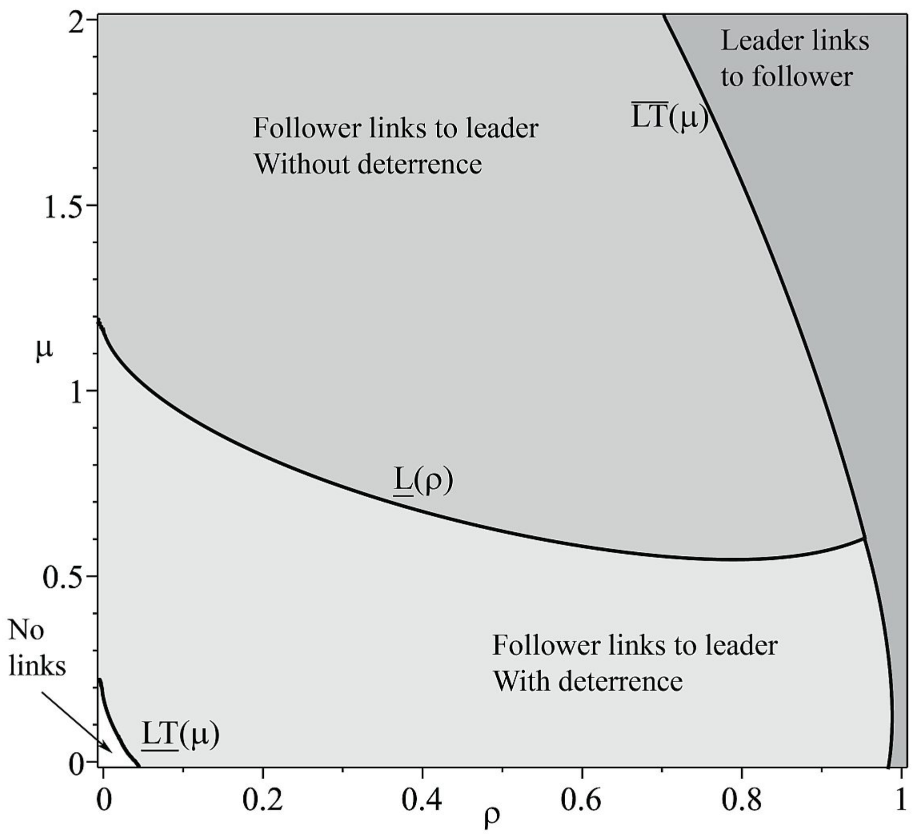


Figure 10: Equilibrium regions with sequential moves.

2. *If $\underline{LT}(\mu) < \rho < \overline{LT}(\mu)$, then the follower links to the target. Furthermore, if $\underline{L}(\rho) \leq \mu$, then*

$$c_1^* = c_T \text{ and } c_2^* = c_S,$$

where $\underline{L}(\rho), c_T$ and c_S are the same as in the basic linking equilibrium. If $\mu < \underline{L}(\rho)$, then the leader produces content above c_T to deter the follower from direct competition, forcing it to link.

3. *If $\overline{LT}(\mu) < \rho$, then the leader links to the follower and*

$$c_1^* = c_S \text{ and } c_2^* = c_T,$$

As expected, the linking equilibrium is the most likely outcome if players move sequentially. Figure 10 depicts the parameter regions where the different types of linking equilibria are feasible. For the most part, the equilibrium is such that the follower links to the target who produces high quality content. In some cases, the source and target content levels are the same as in the simultaneous-move case, and the leader can set the optimal content level expecting the follower to link. However, if μ is below $\underline{L}(\rho)$ and competition is relatively intense, setting the optimal content level is not enough for the leader. The leader needs to increase its content above the (otherwise) optimal link-target content level to deter the follower from competing directly in content, forcing it to link. Note that in this region the linking equilibrium would not be feasible with simultaneously moving players. What makes linking possible with sequential moves is that the follower can expect the leader to stick to its original decision and to not revise its content downwards. It is questionable how credible this is in the online world where links and content levels can change frequently. Although in the majority of the relevant parameter ranges the equilibrium with a link from the follower to the leader is the only possible outcome, there are two other types of equilibria. When both ρ and μ are very small, that is, when the outside option is weak, hence the competition for anchor traffic is tough and when most visitors click through links, the unique equilibrium is no linking. The benefit of linking for the follower is so low that the leader cannot profitably set a high enough content to deter the follower from competing. On the other extreme, when ρ is close to 1 and the link source is able to keep most of the traffic that it attracts using the target's content, the leader will find it profitable to become the source and link to the follower. The follower in this case has no option but to be a link target and set the appropriate (relatively high) content level. Note that although the source's content level is low, it makes higher profits because it profits from most of its visitors.

4 Discrete content choice and mixed strategies

Our basic model with two symmetric nodes results in linking equilibria that are asymmetric with one site linking to the other. Since there are no symmetric linking equilibria in pure strategies it is natural to examine mixed strategies. Unfortunately, mixed strategies can be rather complicated when content decisions are continuous, since in addition to mixing in content decisions, sites can also decide to link with a certain probability, which results in a joint distribution for content and linking. To be able to examine mixed strategies, we resort to a simpler version of the model where content can only be set at two predetermined levels: low and high, denoted by $\underline{c}$ and $\overline{c}$. Since only the fraction of these content levels is relevant we can normalize $\overline{c}$ to 1. Furthermore, we can simplify the cost function since the content decision of a site comes down to the choice between low and high content levels. Therefore, we can focus on the cost difference between these two content levels, denoted by K. Otherwise, this model is identical to our basic model with two nodes. Both sites decide their content levels simultaneously and whether to link to each other or not. We first analyze the equilibria in pure strategies. The payoffs in the no linking cases are the following:

$$\pi_{LL} = \frac{\underline{c}^2}{2\underline{c}+\mu}, \quad \pi_{HH} = \frac{1}{2+\mu} - K, \quad \pi_{LH} = \frac{\underline{c}^2}{\underline{c}+1+\mu}, \quad \pi_{HL} = \frac{1}{\underline{c}+1+\mu} - K,$$

where $\pi_{LL}$and π_{HH} denote the payoffs when both sites choose low (high) and π_{LH} and π_{HL} denote the payoffs when one site choose low and the other high content. Simple analysis of this two-by-two game shows that there are three possible equilibria in pure strategies.

1. When $K > \frac{1}{\underline{c}+1+\mu} - \frac{\underline{c}^2}{2\underline{c}+\mu}$, both sites produce low content levels.

2. When $\frac{1}{\underline{c}+1+\mu} - \frac{\underline{c}^2}{2\underline{c}+\mu} > K > \frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu}$,one site produces low content and the other produces high content.

3. Finally, when $\frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu} > K$, both sites produce high content levels.

That is, costs have to be low enough for sites to invest in high content. For direct comparison with our continuous model, we will focus on the third type of outcome and assume $\frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu} > K$ from this point on while examining linking incentives. Linking is only feasible when the link source has low content and the link target has high content. The source's and the target's payoffs are the following:

$$\pi_S = \frac{\rho\underline{c}}{2+\mu}, \quad \pi_T = \frac{2-\rho}{2+\mu}.$$

In order for linking to be an equilibrium we have to make sure that the source does not have an incentive to deviate to produce high content and sever the link. A deviation is non-profitable iff $\pi_S > \pi_{HH}$. Examining the no linking equilibrium yields the opposite constraint, resulting in the proposition.

Proposition 12. *Assuming* $\frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu} > K$, there are two feasible equilibria

1. *If* $K < \frac{1-\rho\underline{c}}{2+\mu}$*, both sites produce high content levels and do not link to each other*

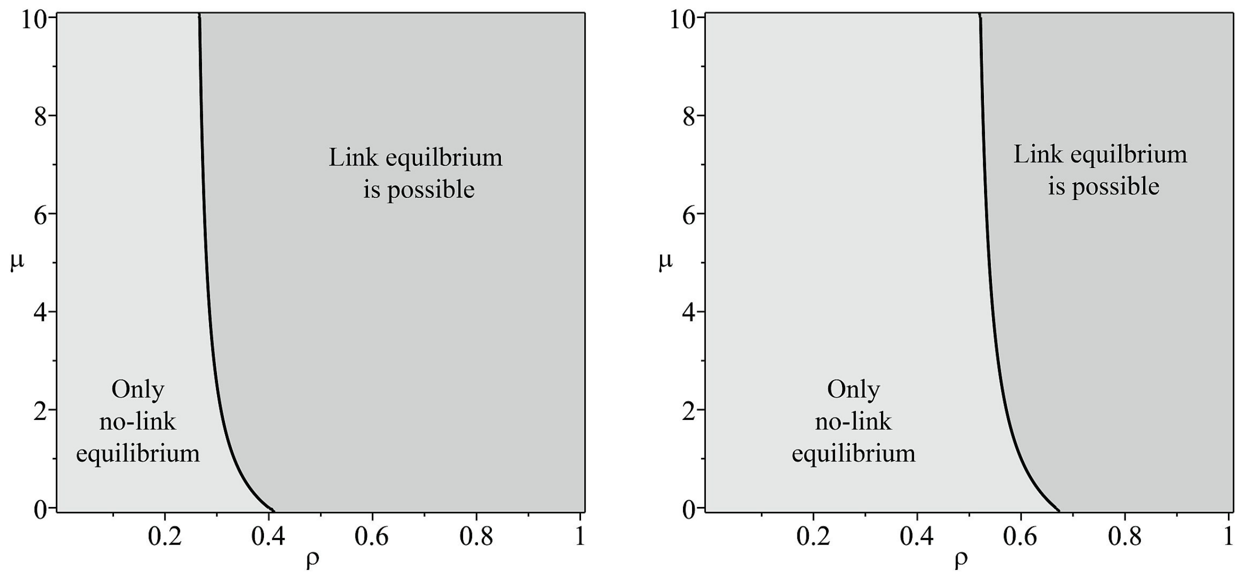
2. *If* $K > \frac{1-\rho\underline{c}}{2+\mu}$*, one site produces low content and links to the other site, which produces high content.*

Note that this result is very similar to our results in the basic model, where linking was not possible when the cost of content production was low, because sites find it profitable to produce high content themselves. If the costs are above a certain level, this changes and one site finds it profitable to produce low content and link to its high-content peer. Depending on the values of ρ and μ the regions where linking is possible can be empty. In fact, linking is possible with a suitable cost if and only if

$$\rho \geq \frac{\underline{c}(2+\mu)}{\underline{c}+1+\mu}.$$

Figure 11 shows that the region where linking is possible exhibits similar features as in the basic model. Just as in the continuous case, linking is only possible when μ is high so that competition between sites is relatively weak. Furthermore, linking in this case requires that ρ is high enough that the link source can capture a high percentage of the extra traffic. Finally, the region where linking is possible expands as the content difference between high and low increases which is consistent with the previous results.

Now that we have examined the discrete case in detail, let us analyze mixed strategy equilibria. In particular, we are going to focus on the parameter region, where the pure strategy equilibrium is asymmetric, that is, where a low quality site links to a high quality site. Thus, we will assume $\frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu} > K > \frac{1-\rho\underline{c}}{2+\mu}$ from now on. Since linking to a high quality site is always better than producing low quality without linking, sites must mix between producing high quality content



(a) $\underline{c} = 0.25$ (b) $\underline{c} = 0.5$

Figure 11: Feasibility of linking equilibrium in the discrete content choice case.

without linking and low quality content with linking. The only outcome that we have not encountered before is if both sites decide to produce low quality content and link to each other, but the payoffs will be simply π_{LL}. Let us determine the symmetric equilibria in mixed strategies and let p_L denote the probability of setting low quality content while linking to the other site. In equilibrium the mixed strategy of one site must make the other site indifferent between the two options, therefore

$$p_L \pi_{LL} + (1 - p_L)\pi_S = p_S \pi_T + (1 - p_S)\pi_{HH}.$$

This leads to the following result.

Proposition 13. *If* $\frac{1}{2+\mu} - \frac{\underline{c}^2}{\underline{c}+1+\mu} > K > \frac{1-\rho\underline{c}}{2+\mu}$, the unique symmetric equilibrium of the game consist of setting low quality content and linking to the other site with probability

$$p_L = \frac{2\underline{c}+\mu}{1-\underline{c}} \cdot \frac{K(2+\mu)+\rho\underline{c}-1}{(2(1-\rho)+\mu)\underline{c}+\mu(1-\rho)},$$

and setting high content without linking with probability $1 - p_L$.

The result is rather intuitive. As the cost of producing high quality content increases, it is more likely that sites will choose to produce low quality content and establish a link. When the cost reaches a critical level the probability reaches 1, but then linking will be worthless as nobody produces high quality content and there is no point in linking. The probability of becoming a link a source also increases in ρ since the link source can capture more of the anchor traffic.

5 Link refusal between two homogeneous sites

Here, we modify the basic homogeneous two player (no aggregator) model and allow link targets to veto (refuse) a link after it has been created by the link source. In contrast to the simultaneous linking and content decisions that we examined in the case without link refusal, in this model sites first make their decisions on content investments, and then they decide whether they want to link to another site. Finally, each link target decides whether it allows or refuses the link.[4]

As one would expect, a site refuses a link from its peer if the link source's content is low relative to the link target. Simple calculations show that a link from site i to site j will be accepted by the target as long as

$$c_i \geq \frac{\rho c_j - \mu(1-\rho)}{2-\rho}.$$

Otherwise, the link will be refused by the target and sites will rely on their own content. When sites are heterogeneous, the above argument implies that targets will refuse links from inefficient competitors. When the competing sites are homogeneous, the following result summarizes how this affects content and linking decisions and how it changes the results of Proposition 1 of the paper..

Proposition 14. *There exist thresholds* $\overline{SCL}(\rho) \geq \underline{CL}(\rho) \geq \underline{L}(\rho)$ *and* $\overline{NCL}(\rho) \geq \overline{NL}(\rho)$ *such that:*

1. *If* $k\mu \leq \overline{NCL}(\rho)$ *then sites do not establish links in equilibrium and* $c_i^* = c_j^* = c_{NL}$.
2. *If* $k\mu \geq \overline{SCL}(\delta\rho)$ *then there are two asymmetric equilibria where one site links to the other and* $c_i^* = c_S \leq c_j^* = c_T$*, which is identical to the case without possible link refusal.*
3. *If* $\overline{SCL}(\rho) \geq \mu \geq \underline{L}(\rho)$ *then there are two asymmetric equilibria where one site links to the other and*
$$c_i^* = \frac{\rho c_T - \mu(1-\rho)}{2-\rho} \geq c_S, \quad c_j^* = c_T$$
4. *There is no equilibrium in pure strategies otherwise.*

Figure 12 illustrates that the results are similar to the case when links cannot be refused but the ability to refuse links makes linking less likely. The parameter region in which linking is feasible shrinks, especially for high values of ρ. This is not surprising, since link targets are hurt the most when link sources are able to retain a high proportion of the traffic they attract. The region of the no-link equilibrium, on the other hand, expands, and the net result is a potential reduction in profits due to the increased competition without links. Nonetheless, for a wide range of parameters, linking is still feasible. If μ is high enough this happens because both the target and the source are better off when linking due to the reduced competition within the content ecosystem and increased audience of link equilibria that is diverted from the competitive outside alternative.

We observe an interesting phenomenon in an intermediate region: when μ is not high enough to provide enough benefits to the target under our standard linking equilibrium, the link source will

[4] Our results do not change if we reverse the order of link creation by the source and the refusal decision by the target as long as these decisions are made after the content investments have been settled.

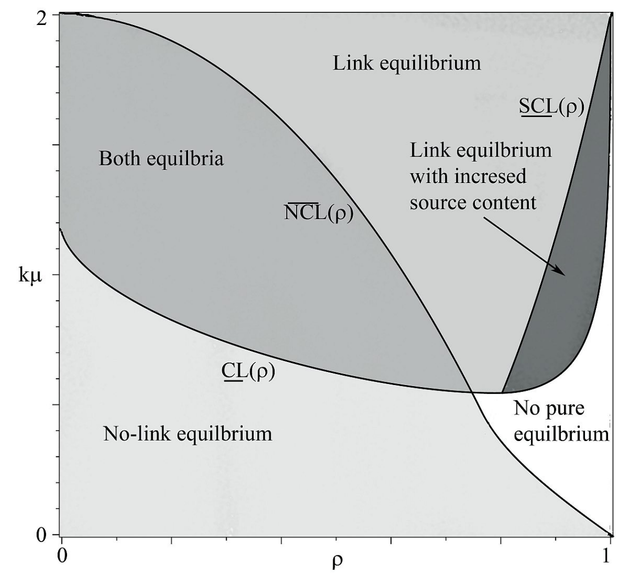


Figure 12: When links have to be approved by the target, the region where linking is sustainable shrinks, whereas the region where there is no linking expands. In the darkest shaded area linking is sustainable, but the link source invest higher than optimal in content to avoid having its link refused.

invest more in content to avoid having its link refused. What is interesting is that in this case the link source produces more content, not so that it attracts more consumers, but so that the target views it as a sufficiently credible competitor with whom it would rather coordinate (i.e. receive a link from) rather than compete head-on.[5] Naturally, this increased content level will be suboptimal given the link, since users will derive all of their utility from the link target anyway, thus the link source's profits will be reduced while the link target's profits will remain unchanged relative to a setting where the target cannot refuse the link. This latter point illustrates how the threat of link refusal hurts one player without necessarily helping the other one (or consumers). In summary, the ability to refuse links may help potential link targets to avoid losses from being linked to, but in many cases can reduce joint profits.

6 Basic model with $\delta < 1$

In the paper we assume that $\delta = 1$, that is, that link sources enjoy the entire benefit of the link target's content. We do so mainly for the sake of an easy presentation. Although assuming $\delta = 1$ may not be realistic relaxing this assumption does not affect our results. For completeness we present the analysis of the basic model with $\delta \leq 1$. Naturally, the part of the analysis concerning the no-linking case is the same as δ only affects players with links. When links are allowed we get the following results:

Proposition 15. *There exist thresholds* $\underline{L}(\delta, \rho), \overline{NL}(\delta, \rho) \in [0, 2]$ *such that:*

[5] This is, therefore, a variant of a hold-up situation.

1. *If* $k\mu \leq \overline{NL}(\delta,\rho)$ *then sites do not establish links in equilibrium and* $c_i^* = c_j^* = c_{NL}$.

2. *If* $k\mu \geq \underline{L}(\delta,\rho)$ *then there are two asymmetric equilibria where one site links to the other and*

$$c_S = \frac{\delta\rho}{k(1+\delta)} \cdot \frac{1+\delta(1-\rho) - 2k\mu + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}}{1+\delta(1-\rho) + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}} \leq$$

$$\leq c_T = \frac{1}{k(1+\delta)} \cdot \frac{1+\delta(1-\rho) - 2k\mu + \sqrt{(1+\delta(1-\rho))^2 + 4(1+\delta(1-\rho))k\mu}}{2}.$$

3. *There is no equilibrium in pure strategies otherwise.*

Proof. There are two possible types of equilibrium with respect to linking: (i) the one where there is no link between the two sites and they invest equally in content (c_{NL}), and (ii) the one where one site invests less in content and links to the other site. As we have already determined the potential equilibria of the first type, we will now identify the candidates for linking equilibria, then check when neither site has an incentive to deviate from a potential equilibrium. When site i links to site j, then its profit becomes

$$\pi_{i\to j} = \frac{\delta c_j \rho}{\delta c_j + c_j + \mu} c_i - \frac{k_i}{2} c_i^2.$$

Comparing these two yields that site i will link to site j iff

$$c_i \leq \frac{\delta\rho c_j (c_j + \mu)}{(1+\delta(1-\rho))c_j + \mu}.$$

Note that the right hand side of the above equation is increasing in c_j and always less than or equal to c_j, yielding that *only the lower quality site will establish a link and only if its quality is sufficiently low relative to its competitor.* Given the above described linking behavior, sites will choose their content investments to maximize profits. Although the site that ends up with a higher content does not consider linking, its profit function changes if its low content competitor decides to link to it:

$$\pi_{j\leftarrow i} = \frac{c_j + \delta(1-\rho)c_j}{\delta c_j + c_j + \mu} c_j - \frac{k_j}{2} c_j^2.$$

Differentiating $\pi_{j\leftarrow i}$ with respect to c_j yields that site j will invest c_T in content if site i links to it (as given in the proposition). Then, differentiating $\pi_{i\to j}$ with respect to c_i yields that site i will invest

$$b_{i\to j}(c_j) = \frac{\delta\rho c_j}{(c_j + c_j\delta + \mu)k}$$

in content if it links to j, yielding the stated $c_i = c_S$ if we plug $c_j = c_T$.

To check whether sites have no incentives to deviate from the potential equilibria, we examine whether the no linking best response would yield higher profits in the linking case and whether the linking best response would yield higher profits in the no-link case. In the first case, the linking equilibrium holds iff

$$\pi_i(b_i(c_T), c_T) \leq \pi_S := \pi_{i\to j}(c_S, c_T).$$

Let $\underline{L}(\delta, \rho)$ denote the value of μ where the above holds with equality when $k = 1$. It is easy to check that the above inequality is invariant to the values of μ and k as long as μk is fixed. Furthermore, it holds for high values of μk, yielding that the linking equilibria exists iff $\mu k \geq \underline{L}(\delta, \rho)$. Similarly, let $\overline{NL}(\delta, \rho)$ denote the value of μk for which $\pi_i(c_{NL}, c_{NL}) = \pi_{i\to j}(b_{i\to j}(c_{NL})_S, c_{NL})$. Sites do not have an incentive to deviate from the no-link equilibrium iff $\mu k \leq \overline{NL}(\delta, \rho)$, completing the proof. □

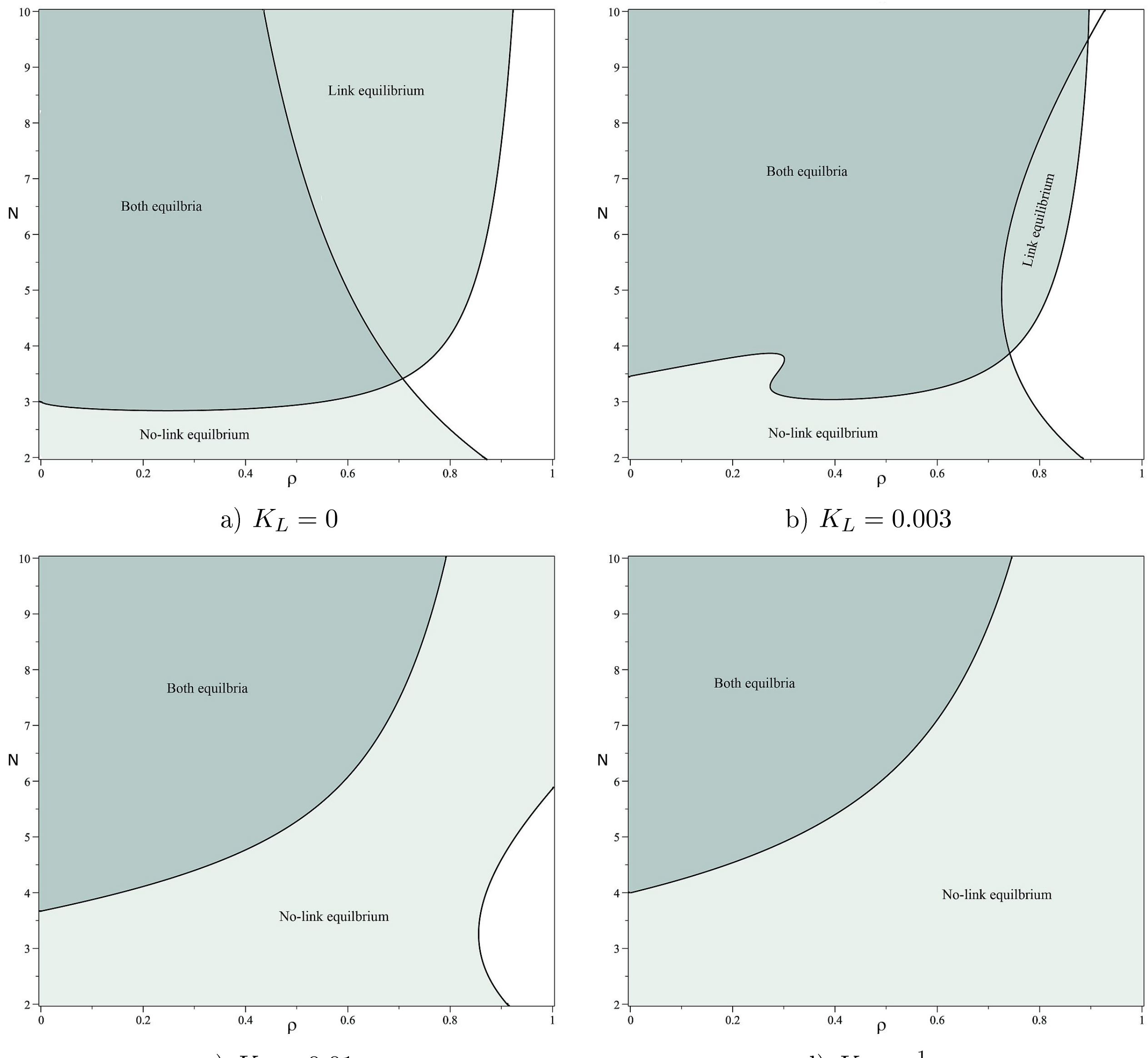
7 Multiple sites: high linking costs and more than one link target

In the paper, we analyze the model with $N > 2$ sites and small K_L>0 costs. We conclude that in most cases there are no links, or if there are, there is a single link target ($N_T = 1$). Here, we provide details about the cases when K_L is not necessarily close to 0 and when $N_T > 1$. We start by discussing the implications of increasing $K_L$on the equilibria with no links or a single link target then we analyze the case of multiple link targets. The existence of an equilibrium without links hinges on the lack of profitable deviations for sites in which they would become link sources and link to their peers. Naturally, as the cost of linking increases such a deviation is even less profitable. Therefore, the parameter region where an equilibrium without links is possible expands as K_L increases. That is, if a no linking equilibrium exists for a certain (ρ, N) pair for K_L,then a no linking equilbrium has to exist for the same (ρ, N) pair for a higher cost $K'_L > K_L$. Indeed, by analyzing the profit in the no linking equilibrium and the profit with a potential deviation (not counting the linking cost), the increment is at most $1/32$, therefore a no linking equilbrium always exists (for any ρ and N) iff $K_L \geq 1/32$. While the no linking equilibrium region expands as $K_L$increases, the opposite is true for the region where a single link target equilibrium is possible. However, such an equilibrium will always be possible with a high enough N, even when K_L is very high (Figure 7 shows the equilibrium regions as $K_L$increases).

Now we move on to examine equilibria with multiple targets. We first focus on $K_L \to 0$, then examine the case of higher costs. As stated in the paper, in equilibrium, we must have

$$c_S = \frac{\rho}{N}, \quad c_T = \frac{(2N_T - 1)(\rho N_T + (1-\rho)N)}{N \cdot N_T^2}.$$

In order to determine the feasibility of such an equilibrium, we have to rule out profitable deviations. The condition for profitable deviations for sites that are link sources are essentially the same when there is one target or more. However, the condition for a link targets' deviation is fundamentally different. When there is only one link target, it can never deviate profitably, since there is no other reasonable quality sites to link to. On the other hand, when there is more than one link target, each link target can consider lowering its content and linking to another link target. The profit that a



a) $K_L = 0$ b) $K_L = 0.003$

c) $K_L = 0.01$ d) $K_L = \frac{1}{32}$

Figure 13: Equilibrium regions as K_L increases.

target makes in the equilibrium candidate is (when its content is c_T)

$$\pi_T^* = \frac{1}{2}\frac{(2N_T-1)(\rho N_T+(1-\rho)N)^2}{N^2 N_T^4}.$$

A deviation from this equilibrium would entail becoming a source and linking to another target. Since other sources are currently linking to this target, its incentives are slightly different from other link sources and the new content level would be slightly higher than c_S:

$$c_S' = \frac{\rho}{N} + \frac{(1-\rho)(N-N_T)}{N\cdot N_T},$$

leading to a payoff of

$$\pi_S' = \frac{1}{2}\frac{((2\rho-1)N_T+(1-\rho)N)^2}{N^2 N_T^2}$$

as $K_L \to 0$. Solving for $\pi_S' = \pi_T$ yields

$$\overline{L}_{N_T}(\rho) = \frac{N_T}{(N_T-1)^2}\cdot\frac{\rho(2N_T-2N_T^2-1)+N_T^2+N_T(1-\rho)\sqrt{2N_T-1}}{1-\rho}.$$

A deviation is always profitable above this threshold, thus, an equilibrium is only sustainable if N_T is below this threshold. Given our equilibrium structure, we naturally need $N \geq 2N_T$, since there has to be at least one source for each link target. Figure 14 shows the limited parameter regions where this type of equilibrium is sustainable. It is not surprising that an equilibrium with multiple link targets is unlikely, since link targets compete with each other and the possibility of giving up the top position is appealing when there is a possibility to link to others. Indeed, the only parameter combinations for which an equilibrium with more than one targets exists are $(N_T=2;\ 4\leq N\leq 14)$, $(N_T=3;\ N=6)$, $(N_T=3;\ N=9)$, $(N_T=4;\ N=8)$, $(N_T=5;\ N=10)$, $(N_T=6;\ N=12)$, $(N_T=7;\ N=14)$.

We continue by examining how the existense of equilibria with multiple targets changes as K_L increases. With a single target we saw that the parameter regions where such an equilibrium is possible shrinks when K_L increases, because link sources have an increased incentive to compete. This is somewhat different when $N_T > 1$. As shown above, the deviation of link sources is not as likely as the deviation of link targets when there is at least one link source and the equilibrium is mostly constrained by this type of deviation. Naturally, as K_L increases link targets have a reduced incentive to become link sources, therefore the constraint is less binding, possibly leading to an expansion of the region where the equilibrium is feasible. However, as K_L increases, the other constraint, the condition on the deviation of link sources becomes binding and reduces the feasible region. That is, an increase in K_L induces two opposing forces on the feasiblity of an equilibrium with multiple targets and it is not clear whether the region is increasing or decreasing. Nevertheless, calculating profits for $\rho = 0$ shows that for high enough K_L and N an equilibrium is feasible with any number of $N_T > 1$ targets. Figure shows how the equilibrium regions where an equilibrium with $N_T = 3$ is feasible changes as K_L increases.

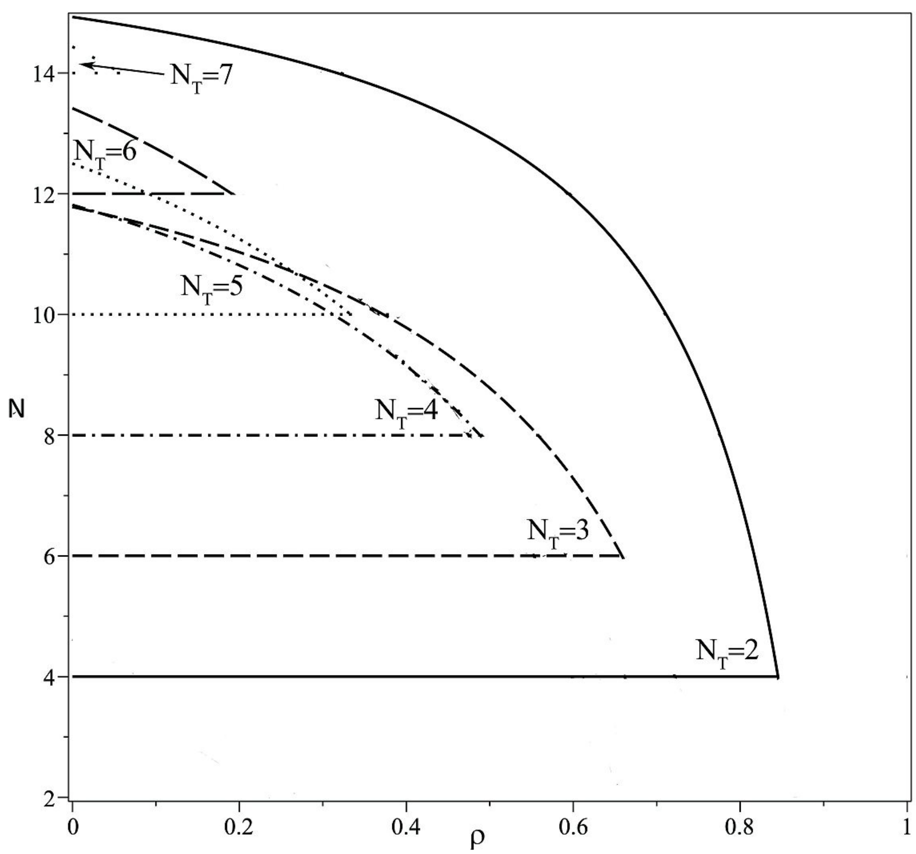
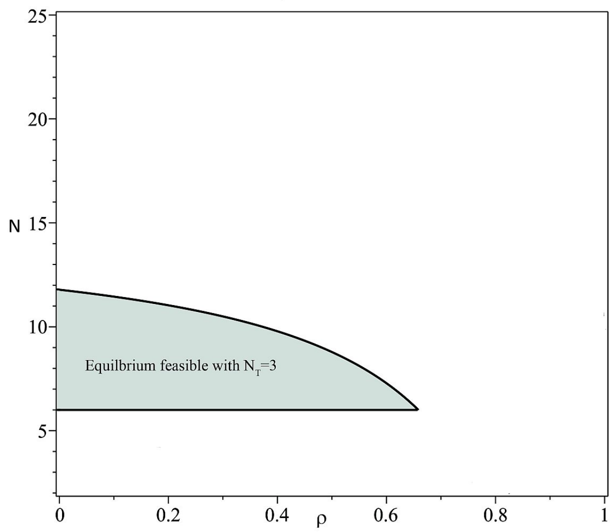
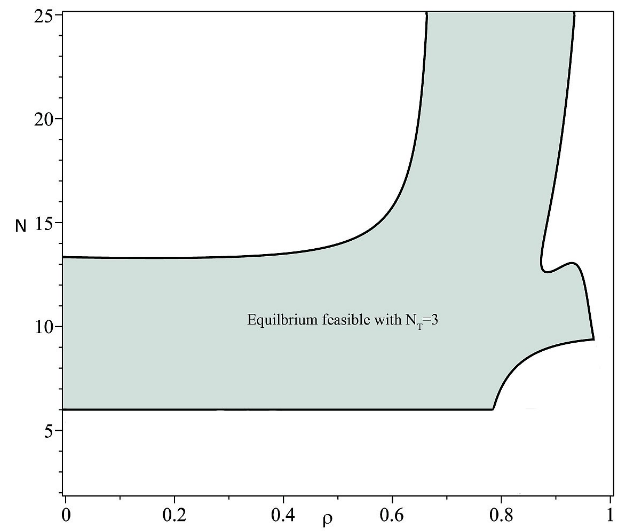


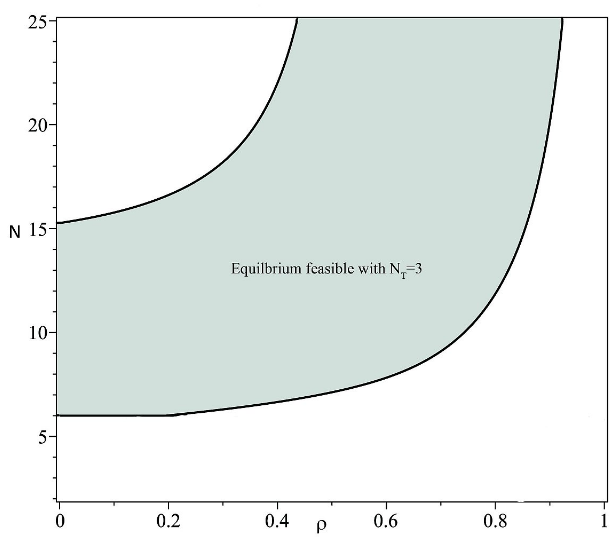
Figure 14: Feasible parameter regions for an equilibrium with $N_T > 1$ link targets.



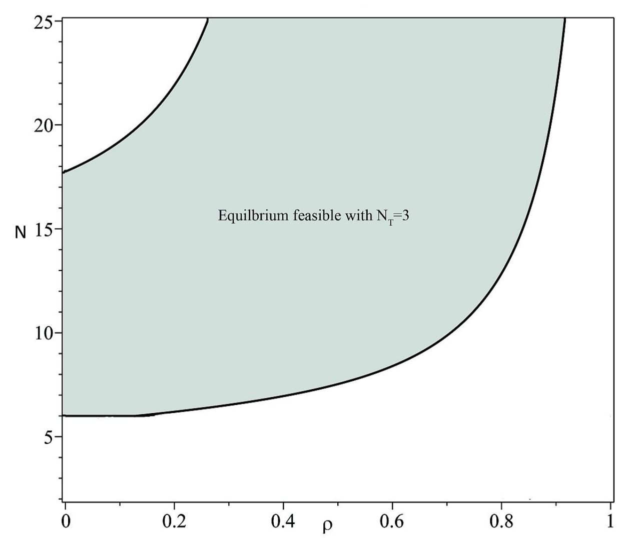
a) $K_L = 0$



b) $K_L = 0.0025$



c) $K_L = 0.005$



d) $K_L = 0.0075$

Figure 15: Equilibrium regions where $N_T = 3$ is feasible as $K_L$increases.