Working Papers
Institute of Mathematical Economics
395
October 2007

Network Formation with Closeness Incentives

Berno Buechel

IMW · Bielefeld University
Postfach 100131
33501 Bielefeld · Germany

email: imw@wiwi.uni-bielefeld.de
http://www.wiwi.uni-bielefeld.de/~imw/Papers/showpaper.php?395
ISSN: 0931-6558

Network Formation with Closeness Incentives

Berno Buechel*

Abstract

We study network formation in a strategic setting where every agent strives for short paths to the other agents. The main parameter of our model is the marginal rate of substitution between network benefits and linking costs. We provide boundaries of stable networks for increasing and decreasing marginal returns.

The formulated model stands in strong relation to the famous connections model (Jackson&Wolinsky '96): we show that for certain parameter values both models induce the same network structures.

* I would like to thank Bastian Westbrock, Dinko Dimitrov, Andreas Brunhart, and Tim Hellmann for insightful comments. I gratefully acknowledge financial support from the international research training group „Economic Behavior and Interaction Models" funded by the German Science Foundation.

Correspondence to Berno Buechel, Institute of Mathematical Economics (IMW), Bielefeld University, D-33501, Bielefeld, Germany. Email: berno@wiwi.uni-bielefeld.de.

1 Introduction

Positions in social networks play a predominant role for economic outcomes. For example, consider a network of R&D collaborations in some technology-based industry. Companies which occupy a very "central" position are considered to better acquire and exploit knowledge what finally fosters their performance (e.g. [15]).

In the field of social network analysis there is a long and rich history of studying benefits of network structures in various contexts. Beyond describing case studies, measures were developed that quantitatively assess the "merit" of certain network positions.[1]

This paper flips perspective by asking how network structures can be affected by agents that strive for beneficial positions. As the motor of each individual's linking behavior we use one of the most customary centrality indices: closeness centrality.[8]

Closeness captures the idea that it is beneficial for an agent to have short paths to many agents in the network. Applications reach from performance of organizational units [17] to status in school classes.[2]

This is not the first paper to use a measure of network centrality as the motor of strategic link formation: Buskens & Van de Rijt [6] study dynamics of intermediary benefits and Rogers [16] models the formation of weighted graphs using an index of social influence. This paper is however the first to work with closeness centrality and to address the open question: *Which network structures emerge when everybody strives for short paths?*

In the next section we will introduce the model. Section 3 provides general results. Section 4 strengthens the assumptions to a linear closeness model and compares this to the connections model introduced by [12].

2 Model

2.1 Framework

<u>Definitions</u>

Let $N = \{1,...,n\}$ be a (finite, fixed) set of **agents/players**, with $n \geq 3$.

A **network/graph** g is a set of unordered pairs, $\{i, j\}$ with $i, j \in N$. This set represents who is linked to whom in a non-directed graph, i.e. $\{i, j\} = ij \in g$ means that player i and player j are linked under g. Let g^N be the set of all subsets of N of size two and G

[1] The underlying assumption of these approaches is that there are some structural features of networks (also called network statistics) that always have an impact on the opportunities or well-being of the agents, be it firms, people, or other units of decision.

[2] Freeman [8] clarifies that closeness measures one dimension of centrality while there are other dimensions, i.e. closeness does not sufficiently capture the intermediary role of some network positions.

be the set of all possible graphs, $G = \{g : g \subseteq g^N\}$. A network can be seen as a (irreflexive, symmetric) binary relation on the player set. It can be represented by a matrix of zeros and ones called adjacency matrix.

Let $L_i(g)$ be the set of links that player i is involved in, that is $L_i(g) = \{ij \in g : j \in N\}$, and $l_i(g)$ its cardinality, called **degree**.

A **path** between two players i and j is a sequence of distinct players $(i_1,...,i_K)$ such that $i_1 = i$, $i_K = j$, and $\{i_k, i_{k+1}\} \in g \quad \forall k \in \{1,...,K-1\}$. The (geodesic) **distance** between two players is the length of their shortest path(s), where the length is the number of links in the sequence. Formally, we can define the distance between two players (d_{ij}) in a graph g by the corresponding adjacency matrix $A(g)$: $d_{ij}(g) := min\{k \in \aleph : A^k(g)_{ij} \geq 1; M\}$. If two players cannot reach one another (there does not exist a path connecting them), we define their distances as M, a number that is bigger than the feasible distances (see section 2.2).

A graph is called **connected**, if there exists a path between any two players in the graph. A set of connected players is called **component**, if they cannot reach agents outside this set. A link is called **critical**, if its deletion increases the number of components in a graph. A graph is called **minimal**, if all links are critical. A **tree** is a connected network that is minimal.

An isolate is a player with degree zero. A pendant is a player with degree one (this structure is called a **loose end**). A **circle** of size K is a sequence of K distinct players $(i_1,...,i_K)$ such that $\{i_k, i_{k+1}\} \in g \quad \forall k \in \{1,...,K\}$, where $i_{K+1} := i_1$.

Game Theoretic Framework

We base our model on a game-theoretic framework introduced by [14], [12], and [1]. Without defining the game explicitly, we take the "shortcut" of working with preferences and directly applying a stability concept.

For each player $i \in N$ a utility function represents his preferences over the set of possible graphs, $u_i : G \to \Re$.

We work with the most basic equilibrium concept due to [12]: a network is considered as "stable" if no link will be added or cut (by one, respectively two, agents). Formally, a network g is pairwise stable (PS) or just **stable** if

(i) $\forall ij \in g$, $u_i(g) \geq u_i(g \setminus ij)$ and $u_j(g) \geq u_j(g \setminus ij)$ and

(ii) $\forall ij \notin g$, $u_i(g \cup ij) > u_i(g) \Rightarrow u_j(g \cup ij) < u_i(g)$.

2.2 Closeness

Motivation of closeness

The unique feature of the model presented here is the utility function, which will be based on closeness.[3] Closeness is based on the idea that an actor prefers networks, in which his distance to the other players is short.

Central positions in terms of closeness are beneficial for multiple reasons:

- *The higher your closeness, the smaller the distance to one arbitrary node in the network.*
  E.g. imagine the joint-work network of Hollywood's actors and directors; where somebody has an idea for a new movie and needs one other artist to pursuit this project. The higher his closeness, the higher the probability to have the desired candidate very few contacts away, such that one can contact him, by just asking a friend.
  Or think of some researcher having a revolutionary idea. In the (co-author) network of this field, people with high closeness are very likely to hear early about it.
- *The higher your closeness, the higher your status.*[4]
  In a friendship network of a school class you can recognize peripheral pupils who are not very popular; and the "central" pupils who seem to identify in a stronger way with the group and enjoy a higher reputation.
  As shown in [5] the structural determinants of trust go beyond direct contacts.
- *The higher the closeness, the better you know the network.*
  Accuracy of knowledge decreases with distances. If you get information third hand and only by one source, you may not have such a balanced view as somebody in the center.
  Short geographic "distances" lead to external economies of scale, e.g. cars in Detroit, Silicon Valley for computers, etc. It is plausible that this argument also holds for closeness in a network.
- *The higher your closeness, the better you can shape the community.*
  Networks with short path lengths facilitate quick diffusion of innovation. Agents with high closeness, therefore, can better spread their ideas.

Definition of closeness

We can generally define closeness such that benefits of an agent *i* gained by network *g*

---

[3] To get an idea of closeness, one can think a node that has the duty to "visit" all other nodes in the network. He travels on the network links and, hence, wants the average distance (for one visit) to be as small as possible.

[4] In some of the examples it is not clear that the causality only goes in this direction. A central position can also be reached *because* of high status or high performance. Our analysis, however, takes this assumption: network positions are beneficial; people form links to optimize their positions.

are decreasing with the (geodesic) distances of *i* to all other agents. To measure closeness there are some more details to look at.

To handle pairs that cannot reach one another you can either restrict the attention to the set of connected graphs, which would be a harsh assumption in a network formation game, or you have to define the distance of unconnected players. Here we define it as M. When *i* and *j* are connected, their distance is in [1, n-1], hence let $M > n-1$.[5]

In the literature on centrality it is standard to normalize an index between 0 and 1. We follow this convention by defining **closeness** of an agent *i* as the following affine transformation of his average distance $\sum_{j\in N} d_{ij}(g) \Big/ (n-1)$:

$$Close_i(g) := \frac{M}{M-1} - \frac{\sum_{j\in N} d_{ij}(g)}{(n-1)(M-1)}.$$

There is another operationalization which is more prominent in the literature: the closeness definition according to Freeman [8]: $FrClose_i(g) := (n-1) \Big/ \sum_{j\in N} d_{ij}(g)$.

The author's trade-off here was that while Freeman's version (inverse distances) is much more customary, our closeness definition (reverse distances) more naturally separates the measurement of a structural feature of a network (network statistic) from its evaluation (as argued in [18]). In the next subsection we will see that this choice does not restrict generality, e.g. if people strive just for Freeman-closeness, this is a special case of our model.

2.3 Model Specification

Our model is based on three major assumptions on individual behavior:

1. *The agents take linking decisions in respect to their degree and their closeness, where closeness is beneficial and links are costly.*
   To get a pure model we exclude all other aspects (that can perturb clear strategic decisions).
   We can think of any decision about adding or cutting links as a proposed exchange of average distance versus degree: You can buy closeness by adding links; you can save costs by passing on closeness.
2. *The utility of a player is composed in an additive way by costs and benefits.*
   This assumption is not very restrictive as utility functions that are not additive separable may be transformed into this form. But it is a very convenient assumption: As the cross-derivatives are zero, the assumption uncouples the effects on utility coming from a change in closeness and a change in degree.
3. *The players are homogeneous in respect to preferences.*

[5] It is often convenient to define $M=n$. In this paper, however, we will keep it as a parameter.

It is an interesting question to ask how networks evolve when players differ in their preferences (see e.g. [9]). However, as applications of our model are very different in nature, we would like to put emphasis on the different contexts that influence everybody's choice, not on the difference between agents (as argued in [4]).

By introducing a (non-decreasing, twice differentiable) benefit function $b:[0,1]\to\Re$ and a (non-decreasing, twice differentiable) cost function $c:[0,n-1]\to\Re$, we can put all assumptions together to what we call the **closeness model**:

the preferences of any player i can be represented by $u_i(g)=b(Close_i(g))-c(l_i(g))$.

Although concave and convex cost functions are reasonable - concave costs represent the combination of fix costs and variable costs; convex costs represent the scarcity of resources (e.g. time) – we will restrict attention to linear cost functions $c^{linear}(g)=\bar{c}l_i(g),\ where\ \bar{c}\in(0,\infty)$. The justification is that a concave or convex cost function would induce similar behavior as the benefit function does when transformed with the inverse function. So these aspects are assumed to be absorbed by the benefit function.[6]

For the benefit function we will distinguish three cases: concave shape, convex shape and linear shape.[7] The first one represents decreasing marginal returns. Formally, $\forall x,x',\Delta>0$ a concave benefit function implies $b(x+\Delta)-b(x)\geq b(x'+\Delta)-b(x')$ whenever $x\leq x'$ (by the mean value theorem). Convexity implies increasing marginal returns: just let $x'\leq x$.

Remark. If you take the following convex benefit function $f(x)=[M-x(M-1)]^{-1}$, then the benefits are equivalent to Freeman-closeness (with linear evaluation), because $f(Close_i(g))=FrClose_i(g)$.

The **marginal costs** are constant at $\bar{c}$ and serve as the parameter for our model. The marginal benefits depend on the network *g* and on the shape of the benefit function. Let $\beta_i^{ij}(g)$ denote the **marginal benefit** that link *ij* (either added or cut) means to player *i* in graph *g*. That is, $\beta_i^{ij}(g):=b(Close_i(g\cup ij))-b(Close_i(g\setminus ij))$.

When players take linking decisions, they compare marginal costs and marginal benefits: in graph *g* player *i* is eager to form a link to *j* ($ij\notin g$) iff $\beta_i^{ij}(g)>\bar{c}$ and *i* wants to cut a link with *k* ($ik\in g$) iff $\beta_i^{ik}(g)<\bar{c}$.[8]

[6] In fact, this assumption restricts preferences to be quasi-linear in degree.
[7] In [10] the role of concave/convex benefits is nicely elaborated. [13] analyzes decreasing marginal returns in a similar model, but with one-sided link formation
[8] When marginal benefits are equal to marginal costs, the player is indifferent. In this case he won't cut the link, respectively does not initiate the new link (but agrees when asked).

3 General Results

This section provides boundaries (thresholds of the parameter) for stable networks in the closeness model and addresses how they can be affected by the curvature of the benefit function.[9]

3.1 Connectedness and loose ends

To have a shorter notation, we substitute two often needed units of closeness:

- $T1 := \frac{1}{(n-1)(M-1)}$. This is the smallest possible change in closeness, as it corresponds to a shift in distance of 1. It occurs when two players, who were at distance two, form a link and only the distance between these two changes, e.g. because they are already directly linked to everybody else.
- $T2 := \frac{1}{(n-1)}$. This is the change in closeness of a player that connects to an isolated node. As his distance shifts from M to 1, his closeness increases by $\frac{M-1}{(n-1)(M-1)} = T2$.

The following results provide two characteristics of all stable networks.

Proposition 3.1. *In a closeness model with linear costs and* ***concave*** *benefits the following holds:*

(i) *If* $\bar{c} < b(1) - b(1-T2)$, *all stable graphs are connected.*

(ii) *If* $\bar{c} > b(T2) - b(0)$, *no stable graph exhibits loose ends.*

Proof. (i) Take any unconnected graph g. Take any player i and let $Close_i(g) =: x$. Linking with somebody of another component leads to a shift in closeness of at least *T2*. Because $x + T2 \le 1$ and $b(\cdot)$ concave, it holds that $b(x+T2) - b(x) \ge b(1) - b(1-T2)$. By assumption the marginal costs are lower, such that i wants to form this link. As in any unconnected graph there exist two players who are not connected, they will alter the network structure, which makes g unstable.

(ii) Take any network g with at least one pendant and let i be his (only) neighbour. Denote $Close_i(g) =: x$. Cutting the link to the pendant means a shift in closeness of *T2*. Because $x \ge T2$ and $b(\cdot)$ concave, it holds that $b(x) - b(x-T2) \ge b(T2) - b(0)$. By assumption the marginal costs are higher. Therefore, i will cut the link, which makes g unstable. ◊

[9] A more comprehensive characterization of the stable networks – which are between these boundaries - is provided in section 4, where we consider one special case.

The intuition behind the result is that the thresholds of (i) and (ii) are just the minimal and the maximal marginal benefit that a link to an isolated node can mean. I.e. the threshold in (ii) is the marginal benefit of a new link in the empty graph $\beta_i^{ij}(g^{empty})$; where the threshold in (i) is the marginal benefit that cutting a link means to the center of a star $\beta_c^{ci}(g^*)$.

If the benefit function is not concave but convex, these two thresholds just switch roles, as stated by the following proposition.

Proposition 3.2. *In a closeness model with linear costs and* ***convex*** *benefits the following holds:*

(i) *if* $\bar{c} < b(T2) - b(0)$ *all stable graphs are connected.*

(ii) *if* $\bar{c} > b(1) - b(1-T2)$ *no stable graph exhibits loose ends.*

The proof is analogue to the proof of proposition 3.1.

Connectedness and non-existence of pendants heavily restrict the candidates for stable networks. This will be exploited in chapter 4.

3.2 Existence

With the assumption of a convex benefit function, there is a very simple – admittedly not a very elegant - way of proving existence of stable graphs.

Proposition 3.3. *In a closeness model with linear costs and* ***convex*** *benefits the following holds: for any parameter value there exists at least one stable network.*

Proof. To show that for any marginal costs $\bar{c} \in (0,\infty)$ there exists a stable network, we take for low costs the complete graph, for high costs the empty network, and in the medium range the star. It is easy to verify that:

- The complete graph is stable iff $\bar{c} \le \beta_i^{ij}(g^N) = b(1) - b(1-T1)$. Remember that *T1* is the shift in closeness when distance increases by 1.
- The empty network is stable iff $\bar{c} \ge \beta_i^{ij}(g^{empty}) = b(T2) - b(0)$.
- A star is stable iff $b(x+T1) - b(x) \le \bar{c} \le \min\{b(1) - b(1-T2), b(x) - b(0)\}$, where $x := \frac{M}{(M-1)} - \frac{2n-3}{(M-1)(n-1)}$ is the closeness of a peripheral player (pendant). To verify the result note that this condition precludes all possible deviations: (a) no peripheral players add a link $\bar{c} \ge b(x+T1) - b(x)$; and (b) the center does not cut a link $\bar{c} \le b(1) - b(1-T2)$; and (c) no peripheral player cuts a link $\bar{c} \le b(x) - b(0)$.

To prove existence for any marginal cost $\bar{c}$, it remains to show that (1) the lower bound of the star is below the upper bound of the complete network and (2) the upper bound of the star is above the lower bound of the empty network (see figure 2).

1) $b(x+T1)-b(x) \leq b(1)-b(1-T1)$ follows from $x+T1 \leq 1$ and convexity of $b(\cdot)$.

2) $b(1)-b(1-T2) \geq b(T2)-b(0)$ follows from convexity of $b(\cdot)$; and
$b(x)-b(0) \geq b(T2)-b(0)$ follows from $b(\cdot)$ increasing and $x \geq T2$. ◊

Figure 1 shows the idea of the proof: For any marginal cost, we can give a trivial example for a pairwise stable network. For concave benefits the thresholds shift such that these trivial graphs do not span the whole parameter space. So in the case of concavity there are two "gaps" for which we could neither prove existence nor non-existence, for all other parameter values, existence is assured.

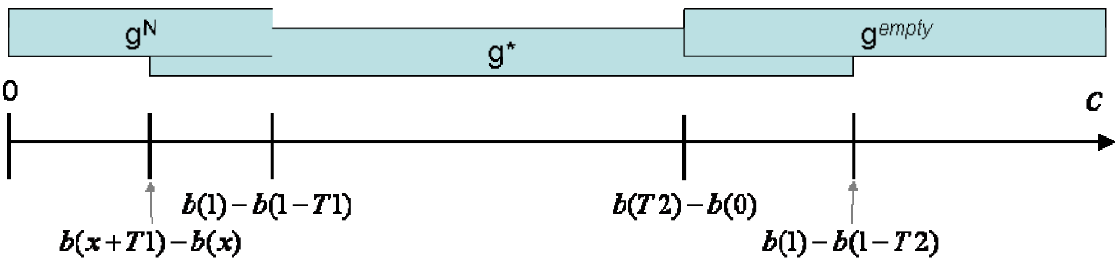

Figure 1: Existence of stable networks for convex benefits.

Remark. Figure 1 also contains the thresholds for proposition 3.2. In the case of concave benefits the two thresholds not only switch positions, but also switch their roles as stated in proposition 3.1.

Besides these trivial examples (empty, complete, star) there are many more stable networks (which will be addressed in section 4).

3.3 Pairwise Nash Stability

The next result addresses the players' incentives to cut links.

Besides pairwise stability there are other equilibrium concepts for network formation models, most of which are refinements of (PS). One of the most used stems from a non-cooperative framework and is called pairwise nash stability (PNS) (see e.g. [2]). We can directly define it by just strengthening condition (i) of (PS): A network g is pairwise nash stable (PNS) if

(i) $\forall i \in N, \forall l \subseteq L_i(g)$, $u_i(g) \geq u_i(g \setminus l)$ and

(ii) $\forall ij \notin g$, $u_i(g \cup ij) > u_i(g) \Rightarrow u_j(g \cup ij) < u_i(g)$.

In the closeness model, (PNS) is not always a proper refinement of (PS).

Proposition 3.4. *In a closeness model with linear costs and* ***concave*** *benefits the set of* *pairwise stable networks* $[PS]$ *and the set of pairwise nash stable networks* $[PNS]$ *coincide.*

One direction of the result follows directly from the definitions: $[PNS] \subseteq [PS]$. The other direction is more intriguing. The complete proof can be found in the appendix. Its main ideas are the following:

[7] show that [PNS] and [PS] coincide, if the utility function $u(\cdot)$ satisfies a property called $\alpha - convexity$ in current links. Moreover, if costs and benefits are additively separable and marginal costs are constant, it is enough to show that the benefit function satisfies $\forall i \in N, \forall g \in G, \forall l \subseteq L_i(g)$,

$$\beta_i^l(g) \geq \sum_{ij \in l} \beta_i^{ij}(g), \qquad (1)$$

where $\beta_i^l(g) := b_i(Close_i(g)) - b_i(Close_i(g \setminus l))$ denotes the marginal benefit that the deletion of the links (in *l*) means to some player *i*.

In essence, the condition says that the deletion of some of player *i*'s links hurts him weakly more than the sequential deletion of these links, one at the time. For constant marginal costs it is intuitive that this is the condition that deviations of cutting more than one link are only utility improving, if deviations of cutting just one link are, which is sufficient for $[PS] = [PNS]$.

To show that condition (1) holds in a closeness model with concave benefits, we need two steps: one step shows that the shift in closeness on the lefthandside of (1) cannot be smaller than the shift in closeness on the righthandside. The other step exploits decreasing marginal returns (which guarantee, roughly speaking, that multiple small reductions of closeness are not evaluated as sever as one big reduction).

The proof of proposition 3.4 clarifies the role of the benefit function for the stability of networks: it is a genuine feature of the model that cutting one link at a time shifts closeness (weakly) less than cutting them at once. The concavity of the benefit function just preserves this feature.

This section provided general results for the closeness model and stressed how they are affected by the curvature of the benefit function. The next section characterizes more explicitly *which networks are stable* by looking at one special case.

4 THE LINEAR CLOSENESS MODEL

In the **linear closeness model**, we assume all players to have a linear cost function and a linear benefit function.[10] Without restriction of generality we take closeness as the numeraire good and represent any player's preferences by $u_i^{linear}(g) = Close_i(g) - \bar{c}l_i(g)$. Note that by taking the id-function as benefit function

[10] As a consequence, the linear closeness model differs from Freeman-closeness (with linear evaluation). But it is equivalent to Freeman-Closeness with a certain concave benefit function.

we mingle in this section what we distinguished before: the closeness of an agent and his benefit derived from closeness.

4.1 Transition Points

The first proposition is a corollary of Prop 3.1 and 3.2, as the linear benefit function is a special case of both, concave and convex benefits functions.

Proposition 4.1. *Let again* $T2 \equiv \frac{1}{(n-1)}$. *In the linear closeness model the following holds:*

(i) *For* $\bar{c} < T2$ *all stable graphs are connected.*

(ii) *For* $\bar{c} > T2$ *no stable graph exhibits loose ends.*

Excluding pendants implies for the stable networks: (a) they cannot be minimal (i.e. a tree); (b) there exists at least one circle if the graph is non-empty; and (c) if the graph is connected, then it must contain at least *n* links.

Typically for the linear closeness model, two thresholds coincide: $b(1)-b(1-T2)=b(T2)-b(0)=T2$. This is also true for the next transition point.

Proposition 4.2. *Let again* $T1 \equiv \frac{1}{(n-1)(M-1)}$. *In the linear closeness model the following holds:*

(i) *For* $\bar{c} < T1$ *the unique stable network is the complete network.*

(ii) $T1 \leq \bar{c} \leq T2$ *a star shaped graph is stable, but not necessarily unique.*

Proof. Remember that *T1* is the shift in closeness when distances shift by 1.
(i) The minimal increase in benefit that a new link can lead to for both its owners is *T*1; because a new link reduces at least the distance to the other player from 2 to 1. So, if costs are strictly lower than this, it follows immediately that nobody wants to cut a link in any graph (stability of complete graph) and any two players, who are not directly linked, will add a link (uniqueness).
(iii) Shown in proof of prop. 3.3. ◊

Costs below *T1* are considered as very small; costs above *T2* are considered as very high. However, *T2* is not necessarily a threshold for uniqueness: There is a third transition point (which can be bigger than *T2*).

Let *T3* be the maximal marginal benefit that a non-critical link can mean to both its owners. We claim that $T3 = \frac{n-1}{4(M-1)}$.[11] *T3* occurs in the line graph, where the pendants form a link. This is just the same marginal benefit that cutting a link in a circle graph means.

Proposition 4.3. *In the linear closeness model the following holds:*

(i) *For $\bar{c} > T3$ every stable graph is minimal or empty.*

(ii) *If $T3 \geq T2$[12], then for $\bar{c} > T3$ the unique stable graph is the empty network.*

Proof. (i) If a non-empty network is not minimal, then there must be at least one non-critical link. By the definition of *T3* networks with such links cannot be stable in this cost range.
(ii) The empty graph is stable because $\bar{c} \geq T2$. For uniqueness note that any non-empty graph must contain either loose ends or circles. By proposition 4.1 (iii) we can exclude all graphs with loose ends for $\bar{c} > T2$. By (i) we can exclude all graphs with circles for $\bar{c} > T3$. ◊

The transition points organize the equilibria in the parameter space. For very small costs and for very high costs, there are only trivial stable networks. In the medium cost range we can find a multitude of stable networks. Figure 3 shows one example of a stable network in the linear closeness model for *n=14*, *M=n* and medium costs.

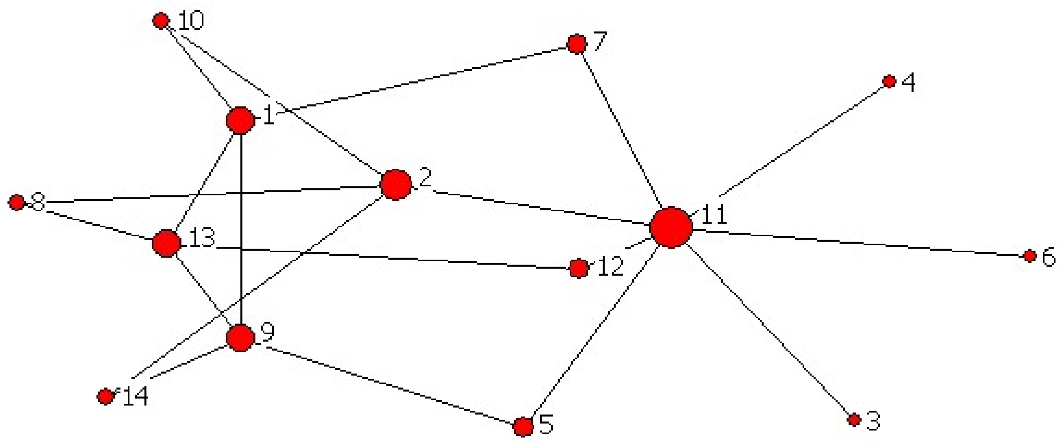

Figure 3: Example of a stable network.[13]

[11] The derivation of the value for *T*3 can be found in the appendix.
[12] In general we will assume that *M* is such that $T3 \geq T2$ holds. We treat the exception of $T3 \leq T2$ in the next subsection as prop. 4.4.
[13] The size and the position of the nodes indicate their closeness.

4.2 Comparison to Connections Model

The linear closeness model is heavily related to the already classic papers of Jackson and Wolinsky [12] and Bala and Goyal [1]. In the famous example of the (symmetric) connections model basically the following benefit is used:

$Connections_i(g) = \sum_{j \in N \setminus i} \delta^{d_{ij}(g)}$ where $\delta \in (0,1)$. So every agent means some worth for *i*, but this diminishes by path length.

[12] already point towards generalizing the model: "we remark that the results presented for the connections model are easily adapted to replace $\delta^{t_{ij}}$ by any non-increasing function $f(t_{ij})$"[14]. The closeness centrality is such an adaption as $b_i(Close_i(g)) = \frac{M}{M-1} - \frac{\sum_{j \in N} d_{ij}(g)}{(n-1)(M-1)}$ is a linearly decreasing function in the sum of distances.

The motivation of both models is quite similar: in both models you gain from having short paths to other nodes.[15] Both models do not consider benefits from having an intermediary position; they are rather about access to resources. But there is also a difference in motivation: In the connections model you benefit from having many nodes close to you; while in the closeness model you benefit from having a small average distance.

Comparison of the stable networks

The linking behavior in the linear closeness model and the (symmetric) connections model can be expected to be quite similar: in both models there is high incentive to link to somebody who is at high distance (or in another component) and there is low incentive to keep links that do not shorten some paths significantly. While the motivation is similar, it turns out that the stable networks are almost identical.

Observe first that propositions 4.1 and 4.2 correspond directly to the results of the connections model, where $T1 \hat{=} \delta - \delta^2$ and $T2 \hat{=} \delta$.

For *n* not too big, a computer can enumerate all networks and check for stability.[16] We did this for *n*=8 with the connections model (taking $\delta = 0.5$ and $\delta = 0.8$), and for the closeness model once with the convex benefit function according to Freeman and once

[14] Where $t_{ij} \hat{=} d_{ij}(g)$.

[15] Borgatti and Everett [3] summarize a group of "Closeness-like measures" and mention the connections model as one of them.

[16] I thank Vincent Buskens for programming the routines to find all the stable networks for the various centrality measures.

taking the linear closeness model (with M=n).[17]

For n=8 there are 12'346 possible isomorphic graphs. In the linear closeness model only 45 of them are stable for some parameter range (larger than 0).[18] As depicted in table 1, those 45 networks are not identical to the 63 stable networks with convex benefit function (Freeman), but overlap to some extent. The stable networks of the linear closeness model and the connections model overlap heavily.

Table 1: Stable networks in the linear closeness model and related models for n=8.

Number of stable networks (for some cost range)	total	also stable in linear closeness model
Freeman closeness	63	29
Connections $\delta = 0.5$	29	26
Connections $\delta = 0.8$	45	45

So we find that, neither the closeness model is a special case of the connections model nor vice versa. With certain specifications they lead to the same network structures.

4.3 Trees

A very special case of the connections model occurs when the decay is very small or zero. Then distances do not matter anymore, crucial is only who you can reach (that is the size of the component). In this context the stable networks are trees.

In the closeness model trees are among the stable networks. We can strengthen this statement for the linear closeness model when M is "big enough"[19] :

Proposition 4.4. *In the linear closeness model for marginal costs in the range* $T3 < \bar{c} < T2$ *the following holds:*

- *(i) all stable networks are trees and*
- *(ii) all trees are stable.*

Proof. (i) Trees are characterized as minimal graphs that are connected. For $\bar{c} < T2$ all

[17] The computer program used slightly different conventions for the treatment of unconnected pairs: In the connections model the distance of unconnected agents is defined as infinity; and for Freeman closeness isolates were normalized to have 0 benefit.

[18] That is: we did not count the networks which are „stable" for only one point in the parameter space, e.g. the networks which are only stable if $\bar{c} = T1$.

[19] Letting $M > \frac{1}{4}(n-1)^2 + 1$ assures that $T2 > T3$.

stable graphs are connected, as shown in Prop. 4.1(i). For $\overline{c} > T3$ all stable graphs are minimal as shown in Prop. 4.3(i).

(ii) A graph is stable, if (a) nobody cuts a link and (b) no two players add a link.

(a) As a tree is minimal, cutting a link leaves two components (unconnected groups of players). The more agents there are in the other component, the higher the loss of benefits. The highest incentive to cut is always given by the neighbor of a pendant, he looses closeness of $1/(n-1) = T2$. By assumption, marginal costs are lower than this (minimal marginal benefit), therefore no agent in a tree will cut a link.

(b) Adding a link to a tree is an addition of a non-critical link (it is a property of trees to be maximally acyclic graphs). For $\overline{c} > T3$ this cannot be favorable for both (by the definition of *T3*). ◊

So in the cost level between *T3* and *T2*, stability of a graph is equivalent to being a tree. Note that many trees are also stable, when costs are below *T3*. Trees are typical outcomes in network formation based on closeness incentives.

5 CONCLUSION

We introduced a network formation model based on closeness centrality. We found very general results on boundaries of stable networks and analyzed the stable networks for a specific case.

The main limitations of the model are its pure assumptions on behavior. In reality people and organizations are not as rational (computing their marginal costs and benefits) when deciding about forming relations; and even if they were, in many cases they would not have enough information to act as purposely. Still, the model provides a benchmark scenario and captures processes that occur whenever actors in a network try to optimize their position in respect to short distances.

To complete the analysis of this model, at least three things need to be done: a characterization of the specific patterns of stable networks, a look on the dynamics of the model, and finally the discussion of efficiency.

For future research on strategic network formation it will be important to systematize the types of incentives and to build models that allow for multiple incentives at the same time.

References

[1] Bala, V., and Goyal, S. (2000): "A Noncooperative Model of Network Formation". In: *Econometrica* **68**, 1181–1229.

[2] Bloch, F., and Jackson, M. (2006): "Equilibrium Definitions in Network Formation Games". In: *International Journal of Game Theory* **34**, 305–318.

[3] Borgatti, S., and Everett, M. (2006): "A Graph-theoretic Perspective on Centrality". In: *Social Networks* **28** (4), 466–484.

[4] Burger, M., and Buskens, V. (2006): "Social Context and Network Formation: Experimental Studies." *ISCORE 234,* Utrecht University.

[5] Buskens, V. (1998): "The social structure of Trust". In: *Social Networks* **20**, 265-289.

[6] Buskens, V., and Van de Rijt, A. (2005): "Dynamics of Networks if Everyone strives for Structural Holes". *ISCORE 227,* Utrecht University.

[7] Calvó-Armengol, A., and Ilkiliç, R. (2004): "Pairwise-Stability and Nash Equilibria in Network Formation Games". *Working paper* [2007-01], Univ. Autònoma de Barcelona.

[8] Freeman, L. (1979): "Centrality in Social Networks: Conceptual Clarification." In: *Social Networks* **1**(3), 215-239.

[9] Galeotti, A., Goyal, S., and Kamphorst, J. (2006): "Network Formation with Heteregeneous Players". In: *Games and Economic Behavior* **54**, 353–372.

[10] Goyal, S., and Joshi, S. (2006): "Unequal Connections." In: *International Journal of Game Theory.* **34** (3), 319-349.

[11] Goyal, S., and Vega-Redondo, F. (2007): "Structural Holes in Social Networks". In: *JET*. forthcoming.

[12] Jackson, M., and Wolinsky, A. (1996): "A Strategic Model of Social and Economic Networks". In: *JET* **71**, 44–74.

[13] Kamphorst, J. (2007): "Network Formation with Decreasing Marginal Benefits of Information". *Mimeo*, Utrecht School of Economics.

[14] Myerson, R.B. (1991): "Game Theory: Analysis of Conflict". Cambridge University Press.

[15] Powell, W.W., Koput, K.W., and Smith-Doerr, L. (1996): "Interorganizational Collaboration and the Locus of Innovation: Networks of Learning in the Biotechnology". In: A*dministrative Science Quarterly*, **41**, 116-145.

[16] Rogers, B. (2004): "A Strategic Model of Network Status". Working paper [2006-01], Caltech.

[17] Song S., Nerur S., Teng J. (2007): "An Exploratory Study on the Roles of Network Structure and Knowledge Processing Orientation in Work Unit Knowledge Management". *ACM SIGMIS Database* **38**(2), 8-26.

[18] Valente, T.W., and Foreman, R.K. (1998): "Integration and Radiality: Measuring the Extent of an Individual's Connectedness and Reachability in a Network". In: *Social Networks* **20,** 89-105.

Appendix

Claim *T3.* *We claim that* $\frac{n-1}{4(M-1)}$ *is the maximal marginal benefit that an inessential link can mean to both its owners.*

The claim is based on the following considerations:

Fix two players *i* and j. Let $\tilde{G} := \{g \in G : 1 < d_{ij}(g) < M\}$ be the set of graphs, where *i* can reach j, without being directly linked. The problem is to find $\arg\max_{g \in \tilde{G}}\_\min\{\beta_i^{ij}(g), \beta_j^{ij}(g)\}$.

To split the problem into cases, let $d_{ij}(g) \equiv d$ and denote $\tilde{G}^d := \{g \in G : d_{ij}(g) = d\}$. Now we solve for each $d \in [2, n-1]$ the problem $\arg\max_{g \in \tilde{G}^d}\_\min\{\beta_i^{ij}(g), \beta_j^{ij}(g)\}$.

Consider the following three notes:

- Because $d_{ij}(g) \equiv d$, there are *d*-1 players on one path between *i* and *j* in any $g \in \tilde{G}^d$. Besides these players there are *n*-*d*-1 players which differ in their roles, denoted by the set $K \subset N$.
- The contribution of any player $k \in K$ to $\beta_i^{ij}(g)$ is at least 0 (e.g. *i*,*k* not connected); and at most $\frac{d-1}{(n-1)(M-1)}$ in the case where the shortest path from *i* to *k* uses *j*.
- Each player $k \in K$ can only contribute either to $\beta_i^{ij}(g)$ or to $\beta_j^{ij}(g)$, but not to both.

Therefore, the maximin occurs in a configuration like in figure "weighted line": All players that are not part of the second shortest path between *i* and *j* are attached either to *i* or to *j* such that *i* respectively *j* is their gatekeeper. Furthermore these disposable nodes are equally distributed between *i* and *j*.

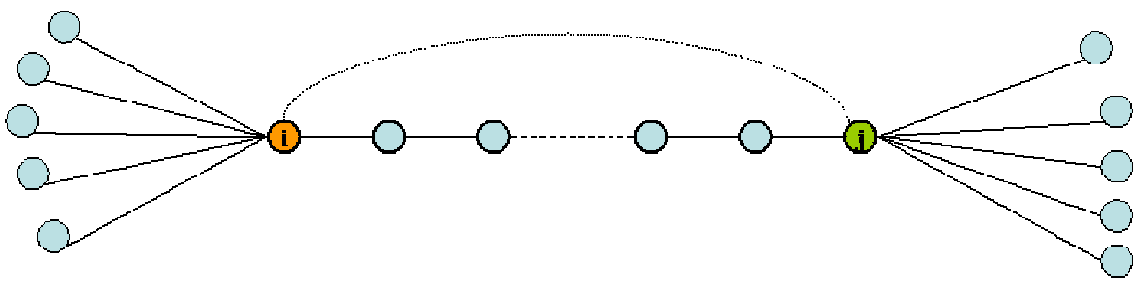

Figure: weighted line. Maximin marginal benefit of a non-critical link.

Claiming that the weighted line of length *d* is the argmax for each problem, we can compute the maximal marginal benefit for each *d*:

It is straight forward to compute that in a weighted line the addition of the link *ij* leads to the following change in the sum of distances for *i* (and *j*)[†]:

$$\sum_{r\in N} d_{ir}(g) - \sum_{r\in N} d_{ir}(g \cup ij) = d - 1 + \frac{1}{4}(d-2)^2 + \frac{1}{2}(d-1)(n-d-1).$$

Now we introduce a function $\hat{\beta} : [n-1] \to \Re$ that relates to any distance *d* the maximal marginal benefit of a link *ij* for *i* (and j):

$\hat{\beta}(d') = [-\frac{1}{4}d'^2 + \frac{1}{2}nd' - \frac{1}{2}n + \frac{1}{4}] \big/ (M-1)(n-1)$. This function attains its maximum at $d = n-1$, which is nothing but a line graph.

By plugging in (and simplifying), we get the result: $\hat{\beta}(n-1) = \dfrac{n-1}{4(M-1)} = T3$.

[†] To use the highest possible value of such a configuration, we took *d* even and *n* odd. For other combinations, the marginal benefit is slightly smaller.

Proposition 3.4. *In a closeness model with linear costs and* ***concave*** *benefits the set of pairwise stable networks* $[PS]$ *and the set of pairwise nash stable networks* $[PNS]$ *coincide.*

Proof.
One direction follows directly from the definitions: $[PNS] \subset [PS]$. For a formal treatment see Bloch&Jackson06. The other direction is more involving.
Calvo-Armengol&Ilkiliç 07 show that [PNS] and [PS] coincide if the utility function $u(\cdot)$ is $\alpha - convex$ in its current links. Moreover, if costs and benefits are additively separable and marginal costs are constant, it is enough to show that the benefit function satisfies $\forall i \in N, \forall g \in G, \forall l \subset L_i(g)$,

$$\beta_i^l(g) \geq \sum\nolimits_{ij \in l} \beta_i^{ij}(g),$$

where $\beta_i^l(g) := b_i(Close_i(g)) - b_i(Close_i(g \setminus l))$ denotes the marginal benefit that the deletion of the links (in *l*) means to some player i. Because of our homogeneity assumption, we can fix a player i without restricting the generality. So we have to show that $\forall g \in G, \forall l \subset L_i(g)$ it holds that

$$b(Close_i(g)) - b(Close_i(g \setminus l)) \geq \sum_{ij \in l} \left[b(Close_i(g)) - b(Close_i(g \setminus ij)) \right]. \qquad (1)$$

In words: the deletion of some of player *i*'s links hurts him weakly more than the sequential deletion of these links, one at the time.

Note the following property of concave functions: for any increasing concave function $f : \Re \to \Re$ it holds that $\forall x, \delta_1, ..., \delta_T, \Delta \in \Re_{++}$,

$$f(x) - f(x - \Delta) \geq \sum_{t=1}^{T} [f(x) - f(x - \delta_t)] \text{ if } \sum_{t=1}^{T} \delta_t \leq \Delta .$$

Let's fix a graph g and a set of links $l \subset L_i(g)$. We substitute $x \equiv Close_i(g)$, $\Delta \equiv Close_i(g) - Close_i(g \setminus l)$, and for $t = 1, ..., T$, $\delta_t \equiv Close_i(g) - Close_i(g \setminus it)$, where every pair in *l* is renamed (in arbitrary order) as *i1, i2,…,iT.*

Using this substitution we learn from the result above that for the (concave and increasing) benefit function $b(\cdot)$

$$b(Close_i(g)) - b(Close_i(g \setminus l)) \geq \sum_{ij \in l} \left[b(Close_i(g)) - b(Close_i(g \setminus ij)) \right] \text{ if}$$

$$\sum_{ij \in l} \left(Close_i(g) - Close_i(g \setminus ij) \right) \leq Close_i(g) - Close_i(g \setminus l) \qquad (2)$$

So (2) is a sufficient condition for (1), for this particular combination of *g* and *l*. As we have to show that the statement (1) holds for any graph $g \in G$ and set of links $\forall l \subset L_i(g)$, we can use the substitution each time and check the sufficient condition

(2). So to proof the result, it remains to show for $\forall g \in G, \forall l \subset L_i(g)$ statement (2) holds.

By using the definition of closeness and after straight forward simplifications, (2) can be rewritten as

$$\sum_{j\in N} d_{ij}(g \setminus l) - \sum_{j\in N} d_{ij}(g) \geq \sum_{ij\in l}\left(\sum_{j\in N} d_{ij}(g \setminus ij) - \sum_{j\in N} d_{ij}(g)\right). \tag{3}$$

We define $\kappa_i^l(g) := \{k \in N : d_{ik}(g) < d_{ik}(g \setminus l)\}$ for $l \subset g$ and for the ease of notation we write it afterwards as κ. We define $\bar{\kappa}_i^l(g) := \bigcup_{ij\in l} \kappa_i^{ij}(g)$ and write it for the ease of notation as $\bar{\kappa}$. κ is the set of players whose distance to *i* increases when the *l* links are cut from g. $\bar{\kappa}$ is the union of players whose distance to i increases when one of the links in *l* is cut.

Now we can transform condition (3) by using the kappa sets (as the summation over all *j* in *N* that are not in any kappa set, cancels out).

$$\sum_{k\in\kappa} d_{ik}(g \setminus l) - \sum_{k\in\kappa} d_{ik}(g) \geq \sum_{ij\in l}\left(\sum_{k\in\kappa_i^{ij}(g)} d_{ik}(g \setminus ij) - \sum_{k\in\kappa_i^{ij}(g)} d_{ik}(g)\right) \tag{4}$$

$$\sum_{k\in\kappa} [d_{ik}(g \setminus l) - d_{ik}(g)] \geq \sum_{ij\in l}\left(\sum_{k\in\kappa_i^{ij}(g)} [d_{ik}(g \setminus ij) - d_{ik}(g)]\right) \tag{4'}$$

Note three properties of the distances in graphs (which are also used in Calvo-Armengol&Ilkiliç 04).

- Note 1: $d_{ik}(g \setminus l) \geq d_{ik}(g \setminus ij) \quad \forall ij \in l$.
- Note 2: $\bar{\kappa} \subset \kappa$.
- Note 3: $\kappa_i^{ij}(g) \cap \kappa_i^{ih}(g) = \varnothing \quad \forall ij \neq ih \in l \subset L_i(g)$.

By note 2 we can split the sum of the LHS in (4'), and we switch the summation on the RHS:

$$\sum_{k\in\bar{\kappa}} [d_{ik}(g \setminus l) - d_{ik}(g)] + \sum_{k\in\kappa\setminus\bar{\kappa}} [d_{ik}(g \setminus l) - d_{ik}(g)] \geq \sum_{k\in\bar{\kappa}}\left(\sum_{ij\in l} [d_{ik}(g \setminus ij) - d_{ik}(g)]\right) \tag{5}$$

$$\sum_{k\in\kappa\setminus\bar{\kappa}} [d_{ik}(g \setminus l) - d_{ik}(g)] \geq \sum_{k\in\bar{\kappa}}\left(\left[\sum_{ij\in l} [d_{ik}(g \setminus ij) - d_{ik}(g)]\right] - \left[d_{ik}(g \setminus l) - d_{ik}(g)\right]\right) \tag{6}$$

By considering note 3 and note 1, the RHS is non-positive. So (6) clearly holds. ◊

Hence, in this model deviations of cutting more than one link are only promising if deviations of cutting just one link are.

Note that the result $[PS]=[PNS]$ together with the definition that $[PNS]=[NS]\cap[PS]$ imply that $[PS]\subset[NS]$.