A network formation game approach to study BitTorrent Tit-for-Tat

Giovanni Neglia[1], Giuseppe Lo Presti[2], Honggang Zhang[3], Don Towsley[4]

[1] D.I.E.E.T., Università degli Studi di Palermo, Italy
INRIA Sophia Antipolis, France, `giovanni.neglia@ieee.org`
[2] IT Dept., C.E.R.N., Switzerland, `giuseppe.lopresti@cern.ch`
[3] Math and Computer Science, Dept. Suffolk University, `hzhang@ieee.org`
[4] Computer Science Dept., University of Massachusetts Amherst,
`towsley@cs.umass.edu`

Abstract. The Tit-for-Tat strategy implemented in BitTorrent (BT) clients is generally considered robust to selfish behaviours. The authors of [1] support this belief studying how Tit-for-Tat can affect selfish peers who are able to set their upload bandwidth. They show that there is a "good" Nash Equilibrium at which each peer uploads at the maximum rate. In this paper we consider a different game where BT clients can change the number of connections to open in order to improve their performance. We study this game using the analytical framework of network formation games [2]. In particular we characterize the set of *pairwise stable networks* the peers can form and how the peers can dynamically reach such configurations. We also evaluate the loss of efficiency peers experience because of their lack of coordination: we find that the loss of efficiency is in general unbounded despite the utilization of the Tit-for-Tat strategy.

1 Introduction

Recently peer-to-peer applications (e.g., BitTorrent [3], Kazaa, eDonkey, and Gnutella [4]) have become very popular. CacheLogic [5] estimates that peer-to-peer generated 60% of all US Internet traffic at the end of 2004 and in particular BitTorrent (BT in what follows), constituted about 30% of Internet backbone traffic in June 2004.

One of the reason of BT success is its ability to enforce cooperation among the peers contrasting the well know problem of free-ride. In fact all the peers interested in a specific file have to announce themselves to a central server, called *tracker*. The tracker maintains the set of active peers, also called the *swarm*, interested in that content and communicates a small random subset of peers from the swarm to each new peer. Peers use this subset to connect to other peers and exchange missing pieces of the file. In general a peer receives many requests for different pieces. In order to decide which requests should be satisfied, the peer uses the Tit-for-Tat strategy: it uploads to the n_u peers (the default value is 4) from which it can download at the highest rate, i.e., its best

uploaders. This strategy is clearly intended to benefit the peers who contribute more to the system.

Tit-for-Tat is generally considered robust to selfish behaviour. To the best of our knowledge, the only analytical support to this belief is in [1]. The authors of [1] study how Tit-for-Tat can affect selfish peers who can change their upload bandwidth in order to try to maximize their downloading rate. We refer to their model as *the rate game*. Under several assumptions, they show that there is a good NE at which each peer uploads at the maximum rate (their model and their results are discussed in more detail in Section 2). However, we observe that BT clients can also change the number of connections to open in order to improve their performance and achieve better performance.

In order to study this aspect we have introduced a new model, which we refer to as *the connection game*. This model captures Tit-for-Tat reciprocation feature by considering that two peers set up a connection between themselves only when they both find it beneficial. We study this game using the analytical framework of network formation games [2]. In Section 3 we characterize the topologies of some *pairwise stable networks* peers can form both in homogeneous scenarios and in heterogenous scenarios (i.e. respectively when all the links have the same or different capacity values). In Section 4 we evaluate the loss of efficiency peers experience because of their lack of coordination: we find that the loss of efficiency is in general unbounded despite the utilization of the Tit-for-Tat strategy. Finally in Section 5 we propose a simple dynamics for this game. We prove that when connection costs are linear functions of the number of links, this dynamics converges to a pairwise stable network. We also quantify by simulations the convergence time and show that as the network size increases the dynamics leads to networks near to the equilibria described in Section 3.

The results about pairwise stable networks in the homogeneous scenario have already been presented in [6]. This paper presents a new result about the heterogeneous scenario and illustrates extensively the results about the dynamics which were just listed in [6]. To the best of our knowledge only three other papers [7–9] use game theory to study the overlay structure arising from the interaction among selfish peers . In these papers peers build an overlay to provide connectivity. The cost of each peer is the sum of the lengths of the shortest paths to all the other peers plus the cost of the links created to connect to the neighbours.

Due to space constraints, proofs are in [10].

2   The Rate Game

Before illustrating the results about Tit-for-Tat we describe in detail the network model considered in [1], because we adopt the same. According to this model every peer has two asymmetric access links to the Internet: one downstream link and one upstream link. Besides it is assumed that bottlenecks can occur only at upstream links. These assumptions are supported by measurement studies (e.g. [11]): most peers in current peer-to-peer networks use cable modem or ADSL
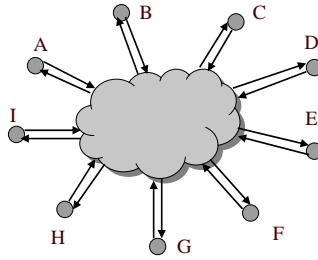
Fig. 1. An example of star network topology.

to get connected to the Internet and usually the data throughput is limited by the "last mile" and the downstream link has higher capacity than the upstream link [11]. Thus, in the star network shown in Figure 1, the Internet cloud can be represented simply as a central node.

A peer r uses its downstream link to get data from other peers. The downstream link of peer r is a "private" link in the sense that this link is only used by peer r itself. On the other hand, the upstream link of peer r is equally shared by all other peers that are downloading files from peer r. We can think of the upstream link of peer r as a "public" link from the point of view of peer r. The model ignore the content dynamics, because it assumes that every peer has potentially interesting data for every other peer and all possible connections can be established.

The authors of [1] study how Tit-for-Tat can affect selfish peers who are able to set their uploading bandwidth in a BT network. Due to Tit-for-Tat reciprocation mechanism, the downloading rate each peer gets is an increasing function of its uploading bandwidth. The authors assume that each peer sets its uploading bandwidth at the minimum level which guarantees them the maximum downloading rate they can achieve, i.e., the downloading rate they would get by uploading at their physical uploading bandwidth[5]. They also assume that the network has a finite number of groups of peers, each of them characterized by a different physical uploading bandwidth and with at least $n_u + 2$ peers. Under these assumptions the authors show that there is a single *good* Nash equilibrium point at which each peer uploads at the maximum rate. Note that in [1], for a given peer, the total number of other peers to set up a connection with is fixed. However, we observe that BT clients can benefit from changing their number of connections (an example is shown in [6]).

---

[5] In reality they need to assume that each peer sets its bandwidth to a value slightly larger than such minimum, otherwise there would be multiple Nash equilibria.

3 The Connection Game

In this section, we first formally introduce our game then we study the network equilibria arising in this game using the analytical framework of network formation games [2].

Assumptions are detailed in the previous section. We refer to peers as players and to connections as links. Let $\mathbf{R} = \{1, 2, \cdots, R\}$ denote the set of players. The strategy of a player i is the set of intended connections player i wants to establish, which is denoted by $s_i = \{s_{i,j} | j \in \mathbf{R} \backslash \{i\}\}$, where $s_{i,j} = 1$ means that player i intends to create a link (open a connection) with player j and $s_{i,j} = 0$ means that player i does not intend to create such a link. With the *Tit-for-Tat strategy*, both players have to agree in order to create a link, hence a link between players i and j is formed if and only if $s_{i,j} = s_{j,i} = 1$. A strategy profile $s = \{s_1, s_2, \cdots, s_R\}$ therefore induces a network $g(s) = \{g_{i,j}, i, j \in \mathbf{R}\}$, where $g_{i,j} = 1$ denotes the existence of link (i, j) and $g_{i,j} = 0$ denotes the absence of link (i, j). Given a network g, we use $g + g_{i,j}$ or $g - g_{i,j}$ to denote the network obtained by adding or severing the link (i, j). We also let $N_i(g) = \{j \in \mathbf{R} : j \neq i, g_{i,j} = 1\}$ be the set of player i's neighbors in graph g, and let $n_i(g) = |N_i(g)|$. A network is symmetric if $n_i(g) = n, \forall i \in \mathbf{R}$, i.e. its topology is a regular graph (all players have the same number of connections).

The payoff or benefit of player i is given by its download rate minus the cost of opening connections: $B_i = G_i - \Phi_i(n_i) = \sum_{j \in N_i(g)} C_j / n_j - \Phi_i(n_i)$, where C_j is the uploading capacity of peer j. We assume that Φ_i is a convex function of n_i (a linear function is a particular case). The marginal benefit for player i to open a new connection with player j is:

$$b_i(n_i(g), n_j(g)) = B_i(g + g_{i,j}) - B_i(g) = \frac{C_j}{n_j(g) + 1} - \Phi_i(n_i(g) + 1) + \Phi_i(n_i(g)).$$

A connection between two players can be set up only when both of them find this connection beneficial. This coordination requirement makes the concept of Nash equilibrium (NE) *partially inadequate*. To address this issue, the idea of NE has been supplemented with the requirement of pairwise stability [12], described below.

Definition 1. *A network g is a pairwise equilibrium network (PEN) if the following conditions hold: 1) there is a NE strategy profile which supports g; 2) for $g_{i,j} = 0$, $B_i(g + g_{i,j}) > B_i(g) \Rightarrow B_j(g + g_{i,j}) < B_j(g)$.*

3.1 Equilibria in Homogeneous Networks

In this section we consider homogeneous networks in which all peers have the same upload capacity and payoff function.

Based on the previous assumptions, our game is the local spillovers game with strategic substitutes properties studied in [13]. Some of the following results (Theorems 1, 2 and 4) can be derived from [13]. Please see Appendix IV in [10] for details.

Theorem 1. *If the number of players is even, a symmetric PEN always exists. Specifically, if $b(0,0) \leq 0$, the empty network is a PEN; if $b(r-2, r-2) \geq 0$, the complete network is a PEN; if $b(k,k) \leq 0 \leq b(k-1, k-1)$, the regular graph with degree k is a PEN. When the previous inequalities are strict, the degree of the PEN is unique.*

Remark. Even when a symmetric network can arise from player interaction according to Theorem 1, the degree of the network is in general different from the default value used in current BitTorrent implementation ($n_u = 4$). This means that the symmetric network created by compliant peers in BitTorrent networks is not in general a PEN for our overlay formation game.

Besides symmetric PENs discussed in the above, we have the following theorem addressing asymmetric PENs.

Theorem 2. *There can be at most one player not connected to any other players in a PEN and the rest of the network is a symmetric network of a unique degree. In asymmetric networks with a single component, if two players with the same number of connections k (i.e. two nodes with the same degree k) are connected to each other, then any two players with fewer number of links than k (or two nodes with lower degrees than k) must be mutually connected.*

Theorem 3. *In a scenario where a unique degree -h- is possible for the symmetric PENs, there can be at most h players with degree smaller than h. Say l the number of players with degree smaller than h, there can be at most $(h-l)l$ players with degree bigger than h, each of them with degree at most $h+l$. If the cost function is linear then there are no players with degree bigger than h.*

Remarks. The two theorems above rule out many possible asymmetric networks, like those with two or more isolated players or interlinked stars[6]. Note that the degree of symmetric PENs h depends only on the cost function $\Phi()$ and the capacity C, and is independent from the number of players R. Because of these theorems the *distance* between a PEN and a symmetric PEN is bounded and becomes less significant as the number of players R increases. Formally:

$$\lim_{R \to \infty} \frac{1}{R} E \left\{ \sum_{i=1}^{R} |n_i(g_{PEN}) - h| \right\} = 0.$$

Similarly the average payoff per player in a PEN converges to that of a symmetric PEN.

The following result shows that players having more connections gain higher payoffs than other players.

Theorem 4. *Let g be a pairwise equilibrium network in which $n_i(g) < n_j(g)$. If $\forall u \in N_i(g)$, $\exists v \in N_j(g)$ s.t. $n_u = n_v$, then $B_i(g) < B_j(g)$.*

---

[6] An interlinked star network has a maximally connected group and a minimally connected group of players. In addition, the maximally connected players are connected to all players while the minimally connected group has links only with the players in the maximally connected set.

Note that if player i's neighborhood is included in player j's neighborhood ($N_i \subset N_j$), the condition "$\forall u \in N_i(g), \exists v \in N_j(g), \text{s.t. } n_u = n_v$" is satisfied.

3.2 Equilibria in Heterogeneous Networks

In this section we consider that peers can have different uploading capacities. Given C_i the capacity of node i, let us indicate k_i a possible node degree in a symmetric PEN when all players have capacity C_i. The following result holds.

Theorem 5. *Under linear costs ($\Phi(n_i) = \alpha n_i$) if in network g each player i has degree k_i, then the network is a PEN.*

Sketch of the proof. We only consider the case $0 < k_i < R-1$. Let us consider $g_{i,j}$. If $g_{i,j} = 1$ then both $b_i(k_i - 1, k_j - 1) \geq 0$ and $b_j(k_j - 1, k_i - 1) \geq 0$ have to be satisfied in order to g be a PEN. $b_i(k_i - 1, k_j - 1) = C_j/k_j - \alpha \geq 0$ because k_j is the degree of a symmetric PEN when all the peers have capacity C_i (see Theorem 1), similarly for $b_j(k_j - 1, k_i - 1)$. If $g_{i,j} = 0$ then we just observe that no player wants to create the link. For example $b_j(k_j, k_i) \leq 0$ again because of the result in Theorem 1.

Remarks. *First*, note that this theorem, differently from Theorem 1 does not state the existence of a PEN. Depending on the values of the capacities, it could be impossible to create a network where all the players open k_i connections (for example if the number of players and the values k_i are odd). *Second*, the result does not hold in general under different cost functions (it is possible to show examples), here it is fundamental that the marginal benefit depends only on the number of connections of the other player. *Third*, in this PEN the distribution of the number of connections in the network mainly reflects the distribution of the capacities. In Section 5 we will show how the PEN selected by the dynamic process we will introduce is "near" to the PEN described in this theorem, and hence the distribution of the number of links resembles the distribution of the capacities in the network.

4 Loss of Efficiency of Symmetric Equilibria

In our game, given the number of players, the number of possible overlays players can create is finite. Hence there is one network g_{opt} with the highest total payoff $\sum_{i \in \mathbf{R}} B_i(g_{opt})$. We define the efficiency loss of a PEN g as the ratio of the highest total payoff over the total payoff of the PEN:

$$L_{eff}(r, C, \Phi) = \frac{\sum_{i \in \mathbf{R}} B_i(g_{opt})}{\sum_{i \in \mathbf{R}} B_i(g)}.$$

We note that L_{eff} depends in general on the number of players, and the upload capacities and cost functions of those players. The following theorem states that L_{eff} is unbounded even for the class of linear connection cost functions ($\Phi(n) =$

αn). Therefore, the price of anarchy (the worst efficiency loss of all NEs) is infinite[7]. Please see Appendix VI in [10] for a detailed proof.

Theorem 6. *For the class of linear connection cost functions, the loss of efficiency is unbounded: given an even number of players and an upload capacity C, $\forall M \in \mathbb{R}, \exists \alpha^* \in \mathbb{R}^+$ s.t. $L_{eff}(r, C, \Phi^*) > M$, where $\Phi^*(n) = \alpha^* n$.*

5 Dynamic Models

We investigate in this section how peers can dynamically reach a PEN. Here we consider linear costs ($\Phi(n_i) = \alpha n_i$). We consider the following dynamic discrete-time process. Starting from an empty network, at each time a player pair (i,j) is randomly chosen. Link (i, j) is created (or kept) if both players find it beneficial. An existing link is removed if at least one of the two players of that link does not find it useful. We are going to show that this dynamic process always reaches a PEN.

Let us introduce some terminology according to [2]. A network g' is *adjacent* to a network g if $g' = g + g_{i,j}$ or $g' = g - g_{i,j}$ for some pair (i, j). A network g' *defeats* another network g if either $g' = g - g_{i,j}$ and $B_i(g') > B_i(g)$, or $g' = g + g_{i,j}$ with $B_i(g') \geq B_i(g)$ and $B_j(g') \geq B_j(g)$ with at least one inequality holding strictly. A network game exhibits *no indifference* if for any two adjacent networks, one defeats the other.

According to this terminology in the dynamic process we described above, the current network is altered if and only if the addition or deletion of a link would defeat the current network. The process leads to an *improving path*, i.e. a sequence of networks $g_1, g_2, ..., g_K$ where each network g_k is defeated by the subsequent (adjacent) network g_{k+1}. There are two kind of improving paths: those exhibiting cycles (which have infinite length) and those terminating with a PEN (called *stable state*). The following lemma (a theorem in [15]) characterizes when there are no cycles and pairwise stable networks exist.

Lemma 1. *Given G the set of all the possible networks g, if there exists a real valued function $w : G \rightarrow \mathbb{R}$ such that "g' defeats g" if and only if "$w(g') > w(g)$ and g' and g are adjacent", then there are no cycles. Conversely, if the network game exhibits no indifference, then there are no cycles only if there exists a function $w : G \rightarrow \mathbb{R}$ such that "g' defeats g" if and only if "$w(g') > w(g)$ and g' and g are adjacent".*

Based on this lemma, we have the following result.

Theorem 7. *If the connection cost function is a linear function $\Phi(n) = \alpha n$, the dynamic process introduced in this section always reaches a PEN.*

---

[7] This is different from what happens for selfish routing, where the price of anarchy is finite, and independent from the network topology for networks in which edge latency does not depend in a highly nonlinear fashion on the edge congestion [14].
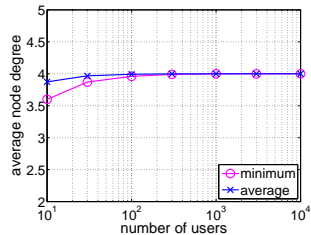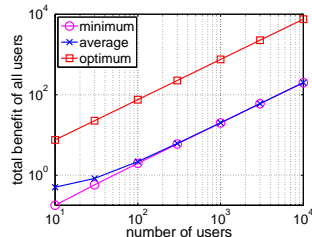
Fig. 2. Average node degree.   **Fig. 3.** Total benefit.

Sketch of the proof. If $h \in \{0, 1, \cdots, R-1\}$ is the degree of a symmetric equilibrium according to Theorem 1 and $b(h, h) < 0$ for $h \neq R-1$, the following function $w : G \to \mathbb{R}$, $w(g) = -\sum_{i=1}^{R} f(n_i)$, where $f(n_i) = h - n_i$, if $h \geq n_i$, and $f(n_i) = R(n_i - h)$ otherwise, satisfies the relation in Lemma 1 for our overlay formation game, hence the dynamic process always reaches a PEN. If $h \neq R-1$ and $b(h, h) = 0$, it is possible to define another function satisfying the relation in Lemma 1. The details of the proof are in Appendix VIII in [10].

5.1 Simulation Results

We present some simulation results. We considered a number of players ranging from 100 to 10000, having the same capacity, and $\alpha = 0.245$, for which the degree of a symmetric PEN is 4. For each setting we simulated 5000 runs of the above dynamic process. Each run terminates with a PEN. For this PEN we denote the average degree over all players as d_{avg}.

Figure 2 shows the minimum and the mean of d_{avg} over all the runs. We see that as R increases both the mean and the minimum converge to 4. This result confirms Theorem 3: as R increases the PENs *converge* to a symmetric one.

In Figure 3, the mean and the minimum of the total benefit are compared with the highest total benefit, which can be directly evaluated from the results in Appendix VII of [10]. This figure shows also the convergence of the payoffs of all PENs to the payoff of the symmetric PEN when R increases.
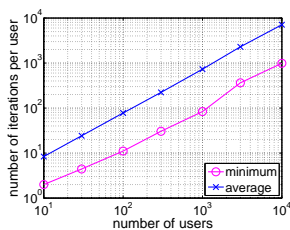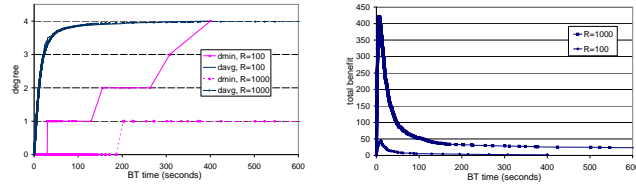

Fig. 4. Number of iterations per peer.

Fig. 5. Convergence to the PEN: the degree.

Fig. 6. Convergence to the PEN: the benefit.

In addition, we present the number of iterations per peer in Figure 4. We observe that the average number of iterations to reach a PEN is of the order of R^2 and hence the number of iteration per peer is of the order of R. Let us consider this number of iterations in the context of BitTorrent (BT) [3]. Each peer in a BT network tries to replace an existing connection with a new, better connection every 10 seconds. All peers do such replacement simultaneously, unlike the sequential replacement in our simulations. So R^2 iterations in our simulations corresponds to $10R$ seconds in a BT network. For a population of 100 peers, the time needed to reach a PEN is of the order of 17 minutes, which is faster than the typical average time between changes in the population of peers (due to arrivals or departures). Figure 5 shows how the average and the minimum degrees change during two simulation runs respectively for $R = 100$ and for $R = 1000$. The initial values are equal to 0 and converge to 4. The time scale represents time in a BT network; namely, R iterations are represented by 10s. We can observe that: 1) with this time scale the evolution of the average degree seems independent from the number of players; 2) the network converges quite rapidly to the PEN. In particular, the average degree reaches 3.8, i.e. 95% of the final value, after less than 80 seconds in both cases, or, equivalently, after less than 800 iterations for $R = 100$ and less than 8000 for $R = 1000$. Figure 6 shows the time evolution of the process as regards the total benefit. We note that for both runs, as the process begins the total benefit grows because of the high benefit of the initial connections, while it falls down to the expected value when the network approaches the equilibrium.

Finally, we considered heterogeneous scenarios where players have different upload capacities. The simulations show that at the final equilibrium almost all the players open a number of connections equal to that indicated by Theorem 5. For example we considered a network with 1000 players and $\alpha = 0.245$, where 50% of the nodes have capacity equal to 1, 30% have capacity equal to 2, and 20% have capacity equal to 4. The corresponding degrees in symmetric networks (k_i in Section 3.2) are respectively 4, 8, 16. In the PEN the fraction of nodes having a degree smaller than the corresponding k_i is on average equal to 0.15% of the total number of nodes, hence the degree distribution closely resembles the capacity distribution.

6 Conclusions

We studied the Tit-for-Tat strategy (built in BitTorrent [3]) through network formation games framework. We proved the existence of equilibrium overlays, and demonstrated the convergence of a simple game dynamics. Although the general belief is that the Tit-for-Tat can prevent selfish behavior, we showed that it can still lead to an unbounded loss of efficiency.

7 Acknowledgements

References

1. Qiu, D., Srikant, R.: Modeling and performance analysis of bittorrent-like peer-to-peer networks. In: Proc. of SIGCOMM '04. (2004) 367–378
2. Jackson, M.O.: A survey of models of network formation: Stability and efficiency. In: Group Formation in Economics: Networks, Clubs and Coalitions, New York, Cambridge University Press (2004)
3. Cohen, B.: Bittorrent, http://www.bittorrent.com
4. Wikipedia: http://en.wikipedia.org/wiki/peer-to-peer
5. CacheLogic: http://www.cachelogic.com
6. Zhang, H., Neglia, G., Towsley, D., LoPresti, G.: On Unstructured File Sharing Networks. In: Proc. of IEEE INFOCOM '07. (2007)
7. Fabrikant, A., Luthra, A., Maneva, E., Papadimitriou, C.H., Shenker, S.: On a network creation game. In: Proc. of the 22nd annual symposium on Principles of distributed computing, New York, NY, USA, ACM Press (2003) 347–351
8. Chun, B., Fonseca, R., Stoica, I., Kubiatowicz, J.: Characterizing selfishly constructed overlay networks. In: Proc. of IEEE INFOCOM'04, Hong Kong
9. Corbo, J., Parkes, D.C.: The price of selfish behavior in bilateral network formation. In: Proc. 24rd ACM Symp. on Principles of Distributed Computing (PODC'05), Las Vegas, Nevada, USA (2005) 99–107
10. Zhang, H., Neglia, G., Towsley, D., LoPresti, G.: On Unstructured File Sharing Networks. Technical Report 06-40, UMass (2006) `http://www-sop.inria.fr/maestro/personnel/Giovanni.Neglia/publications/%Zhang06_UFS.pdf`.
11. Saroiu, S., Gummadi, P., Gribble, S.: A measurement study of peer-to-peer file sharing systems. In: Proceedings of Multimedia Computing and Networking. (2002)
12. Jackson, M., Wolinsky, A.: A strategic model of economic and social networks. Journal of Economic Theory **71**(1) (1996) 44–74
13. Goyal, S., Joshi, S.: Unequal connections. Forthcoming in International Journal of Game Theory (2006)
14. Roughgarden, T.: The price of anarchy is independent of the network topology. In: ACM Symposium on Theory of Computing. (2002)
15. Jackson, M., Watts, A.: The existence of pairwise stablenetworks. Seoul Journal of Economics **14(3)** (2001)