On a Network Creation Game

Alex Fabrikant[*][†]
alexf@cs.berkeley.edu

Ankur Luthra[*]
ankurl@cs.berkeley.edu

Elitza Maneva[*]
elitza@cs.berkeley.edu

Christos H. Papadimitriou[*‡]
christos@cs.berkeley.edu

Scott Shenker[§]
shenker@icsi.berkeley.edu

ABSTRACT

We introduce a novel game that models the creation of Internet-like networks by selfish node-agents without central design or coordination. Nodes pay for the links that they establish, and benefit from short paths to all destinations. We study the Nash equilibria of this game, and prove results suggesting that the "price of anarchy" [4] in this context (the relative cost of the lack of coordination) may be modest. Several interesting extensions are suggested.

Keywords

Distributed network design, game-theoretic models, Nash equilibria, price of anarchy

1. INTRODUCTION

The Internet is the first computational artifact that was not designed by one economic agent, but emerged from the distributed, uncoordinated, spontaneous interaction (and selfish pursuits) of many. Today's Internet consists of over 12,000 subnetworks ("autonomous systems"), of different sizes, engaged in various, and varying over time, degrees of competition and collaboration.

The Internet is also the first object studied by computer scientists that must be approached with humility and puzzlement, and studied by measurement, experiments, and the development of models and falsifiable theories — very much like the cell, the universe, the brain, and the market.

Because the Internet is the product of — and an arena for — the interaction of many economic agents, it may not be optimized in any conventional sense. How costly is this lack of coordination? In Theoretical Computer Science, we have in the past confronted such questions by developing research frameworks based on *ratios:* the *approximation ratio* measures the cost of the lack of tractability; the *competitive ratio*, the cost of information (often in a distributed setting). In [4], it was proposed that the ratio of the social costs of the worst-case Nash equilibrium and the social optimum, the so-called *price of anarchy*, may be an informative measure of the lack of coordination in situations in which agents interact by pursuing each their own interest.

The concept of Nash equilibrium is in some sense the analog of centralized optimal design in the context of multiple distributed selfish agents. But it is not without its conceptual and practical problems. The concept is declarative (as opposed to algorithmic), providing no guidance on how it can be reached (besides some weak and unconvincing convergence results); it is also delightfully and intriguingly non-deterministic (due to the existence of multiple Nash equilibria), something explored and delimited in the current line of work.

Since [4], there has been much progress in understanding the price of anarchy in more and more general situations, in which individual users choose routes selfishly, and suffer from the created congestion [5, 7, 1, 6]. In the Internet, however, users do not choose routes; the situation is much more complex. Routes are chosen by the interaction of packets with routers, users adjust their usage to the resulting congestion, while autonomous systems add bandwidth and hardware to the resulting hot spots. Can this much more complex situation be illuminated as successfully as the congestion game? *What is the price (in both hardware costs and quality of service) of the Internet's open architecture?*

In this paper we propose a simple game-theoretic model of network creation. The agents are nodes, and their strategy choices create an undirected graph. Each node chooses a (possibly empty) subset of the other nodes, and lays down edges to them. The edges are undirected, in that, once installed, they can be used in both directions, independently of which node paid for the installation. The union of these sets of edges is the resulting graph. (The union may not be disjoint; that is, it may be that an edge is paid for by both

of its endpoints. However, this obviouly will never be the case at equilibrium.)

The cost to each node of such a combination of choices has two components: the total cost of the edges laid down by this node (the number of edges times a constant $\alpha > 0$, the only parameter in this model), plus the sum of the distances from the node to all others. Our game tries to capture aspects of the Internet related to autonomous systems peering and otherwise agreeing to communicate. Each such agreement may be costly, but pays off in quality of service improvements. That is, our model takes into account both hardware costs and quality of service costs; however, for the latter it ignores congestion, the focus of much other work.

What are the (pure) Nash equilibria of this game, and how do they compare to the social optimum (the combination of strategies with smallest sum of costs)? (It follows easily from our results that the game does have pure equilibria.)

For small values of α (under 2) the situation is fairly straightforward: the social optimum is the clique, and the star is the worst Nash equilibrium. The worst-case "price of anarchy" is then $\frac{4}{3}$ (see Section 2).

For $\alpha > 2$, the plot thickens. The star is always the social optimum (Section 2), and there is a lower bound of $3 - \epsilon$ for the price of anarchy (Theorem 2). As for upper bounds, we prove that the price of anarchy is $O(\sqrt{\alpha})$ (Theorem 1). This is the best general upper bound that we have.

In our constructions, as well as our experiments, we have almost always come up with Nash equilibria that are *trees*. The only known exception that remains an equilibrium for $\alpha > 2$ (and even then, only for $\alpha \leq 4$), the Petersen graph, is a *transient*[1] equilibrium. For $\alpha \leq 2$, experiments yield a wide variety of non-tree equilibria, but Theorem 1 guarantees that the price of anarchy of all of these is at most constant. (Incidentally, our experiments are hindered by the fact that computing the best response in this game is NP-hard, Proposition 1.)

These considerations have led us to formulate the *tree conjecture*, stating that, for α above some constant A, all Nash equilibria are trees. Finally, we show (Theorem 3) that any tree Nash equilibrium is less than 5 times more costly than the social optimum — hence, under the tree conjecture, the price of anarchy (for non-transient equilibria) is at most a constant, dependent only on A.

Despite its simplicity, our model may be a useful analytical tool for shedding light on the complex processes that create the topology and other reality of the Internet. For example, our analysis can be seen as justifying the prominence of the star and the clique as important primitives in the study of

___

[1] By *transient*, we mean a weak Nash equilibrium (i.e. one in which at least one player can change his strategy without affecting his payoff), from which there exists a sequence of single-player strategy changes which do not alter the changer's payoff leading eventually to a non-equilibrium position. For the Petersen graph, there are several such chains, all leading to positions from which all further chains of single-player changes strictly beneficial to the changer lead to a tree.

the Internet topology. We also outline in the last section several extensions that may make the model a little more realistic, while being still tractable in interesting ways.

The Model

The game we consider has players $\{0, 1, \ldots, n - 1\}$ — we denote this set by $[n]$. The strategy space of player i is the set $S_i = 2^{[n] - \{i\}}$. Given a combination of strategies $s = (s_0, \ldots, s_{n-1}) \in S_0 \times \ldots \times S_{n-1}$, we consider the graph $G[s]$, the underlying undirected graph of $G_0[s] = ([n], \cup_{i=0}^{n-1}(\{i\} \times s_i))$. The cost incurred by player i under s is defined to be

$$c_i(s) = \alpha \cdot |s_i| + \sum_{j=0}^{n-1} d_{G[s]}(i, j),$$

where $d_{G[s]}(i, j)$ is the distance between nodes i and j in the graph $G[s]$.

A (pure) Nash equilibrium in this game is an s such that, for each player i, and for all s' that differ from s only in the ith component, $c_i(s) \geq c_i(s')$.

One may try to discover Nash equilibria by starting from one strategy combination and repeatedly replacing a player's strategy by its best response. The following is cautionary in this regard:

PROPOSITION 1. *It is NP-hard, given $s \in S_0 \times \ldots \times S_{n-1}$ and $i \in [n]$, to compute the best response of i.*

Sketch: Reduction from DOMINATING SET.

Player i is given the configuration of the rest of the graph, including the edges coming in, and has to pick a subset of vertices such that when the edges to them are built, c_i is minimized. For any $1 < \alpha < 2$, if there are no incoming edges, the strategy is a dominating set for the rest of the graph, since the diameter of G must be at most 2, and making more than the minimum number of links would only improve the distance term of c_i by 1. Hence the costs are minimized (and the utility maximized) when the size of the subset is minimized.

∎

2. BASIC RESULTS

The social cost is simply the sum of all players' costs, which, for any situation where no connection is paid for by both endpoints (a constraint satisfied by all Nash equilibria), is:

$$C(G) = \sum_i c_i = \alpha|E| + \sum_{i,j} d_G(i, j).$$

Since every pair of vertices that is not connected by an edge is at least distance 2 apart, the following is a lower bound for the social cost:

$$
\begin{aligned}
C(G) \quad &\geq \alpha|E| + 2|E| + 2(n(n-1) - 2|E|) \\
&= 2n(n-1) + (\alpha - 2)|E|.
\end{aligned}
\tag{1}
$$

This bound is achieved by any graph of diameter at most 2.

When $\alpha < 1$, by Eq. 1, the social optimum is achieved when $|E|$ is maximum. Therefore it is the complete graph. Any Nash equilibrium has to be[2] of diameter 1, which implies that the complete graph is also the only Nash equilibrium.

When $1 \leq \alpha < 2$, the social optimum is still achieved for the complete graph (even though it is no longer a Nash equilibrium). Any Nash equilibrium is of diameter at most 2, so the social cost is exactly equal to that in equation 1. It is worst when $|E|$ is minimum, which is $n - 1$ for a connected graph. Thus the worst Nash equilibrium is the star. The price of anarchy is then:

$$\frac{C(\text{star})}{C(K_n)} = \frac{(n-1)(\alpha - 2 + 2n)}{n(n-1)(\frac{\alpha-2}{2} + 2)}$$
$$= \frac{4}{2 + \alpha} - \frac{4 - 2\alpha}{n(2 + \alpha)}$$
$$< \frac{4}{2 + \alpha} \leq \frac{4}{3}$$

When $\alpha \geq 2$, the social optimum is achieved for minimum $|E|$, so it's the star. The star is also a Nash equilibrium, but there may be worse Nash equilibria.

3. AN UPPER BOUND

Observe that if $\alpha > n^2$ the Nash equilibrium is a tree, because, unless the distance to a node is infinity, a player has no interest in building an edge. In that case, the price of anarchy is trivially $O(1)$.

THEOREM 1. *For $\alpha < n^2$ the price of anarchy for our model is $O(\sqrt{\alpha})$.*

PROOF. The price of anarchy is:

$$\rho(G) = \Theta\left(\frac{\alpha|E| + \sum_{i,j} d_G(i,j)}{\alpha n + n^2}\right)$$

Notice that $d_G(i,j) < 2\sqrt{\alpha}$ for every i and j, since otherwise i would connect to j to make itself closer to all nodes more than half way to j along the shortest path from i to j. Therefore it suffices to prove that $|E| = O(\frac{n^2}{\sqrt{\alpha}})$.

Consider the edges out of vertex v: $e_1, e_2, \ldots, e_l$. For any edge we will count vertices u for which (v, u) is not in the graph. In other words, we will associate several non-edges with each edge. Ideally, we want the ratio of the number of edges to the number of non-edges to be $1 : \sqrt{\alpha}$.

Let $T_i = \{u \in V :$ the shortest path from v to u goes through $e_i \}$. We ensure that T_i are disjoint by considering a canonical shortest path for each vertex. Before edge e_i was built, the alternative shortest path from v to $u \in T_i$ was either

_____

[2]Note that here, and at several points throughout, we rely on the fact that any Nash equilibrium cannot be missing edges whose addition would reduce the sum inter-node distances by more than α, which follows immediately from the definition of the cost function.

infinity or $< 2 * \text{diam}(G) < 4\sqrt{\alpha}$ (where $\text{diam}(G)$ denotes the diameter of G). We consider these two cases separately.

If T_i and v are connected in $G' = (V, E - e_i)$, then for every $u \in T_i$

$$d_{G'}(v, u) - d_G(v, u) < 4\sqrt{\alpha}.$$

The total improvement is $\sum_{u \in T_i} (d_{G'}(v, u) - d_G(v, u)) \geq \alpha$. This implies that $|T_i| = \Omega(\sqrt{\alpha})$. So we found $\Omega(\sqrt{\alpha})$ vertices, such that (u, v) is not an edge. Such non-edges will be counted at most twice (from both sides).

If T_i and v are not connected in G' then G' has two components and we can count $|T_i| - 1 + |V - T_i| - 1 = n - 2 = \Omega(\sqrt{\alpha})$ non-edges – those incident on v or w and the other component, where $e_i = (v, w)$. These are also counted at most twice.

This completes the proof that the number of edges is $O(\frac{1}{\sqrt{\alpha}})$ of the total number of vertex pairs, and the proof of the theorem. $\square$

4. A LOWER BOUND

THEOREM 2. *For any $\epsilon > 0$, there exists a Nash equilibrium of the network game with the price of anarchy greater than $3 - \epsilon$.*

PROOF. For any $k \geq 4$, $d \geq 2$, consider the family of complete k-ary directed trees of depth d, $T_{k,d}$, with all edges going from parent nodes to their children. Let n be the number of nodes in $T_{k,d}$, and set, with foresight, $\alpha = (d - 1)n$.

LEMMA 1. *$T_{k,d}$ is a Nash equilibrium of the network game.*

PROOF. Every non-leaf node i must link at least once to the subtree of each of its children; else the graph becomes disconnected, which would carry an infinite penalty for c_i. If only one link is made to a particular child's subtree (i.e. the set of the child's descendants, including the child itself), it is clearly optimal for it to link to the child directly — linking to a grandchild's subtree brings the node closer to that subtree but further away from the child and at least three other grandchildren's subtrees, thus strictly increasing the distance term. This makes the total contribution of that subtree to the distance term of c_i be at most $(d - 1)n/k$, since the subtree contains at most n/k nodes, and they are all at most $d - 1$ hops from the node we're examining. Since creating 2 or more links to that subtree would require an additional cost of $(d - 1)n$, but would not increase the distance from our node to any other by more than $d - 1$, that will not happen.

Every non-root node i, at depth $\delta \geq 1$, is already connected to all nodes *outside* its subtree via a link from its parent. Note that (1) if such a node links to the root, it will become closer to every other node by no more than $\delta - 1$, for a total possible savings of no more than $(d - 1)(n - 1)$; and (2) if such a node links to a child of its parent (its sibling), it will become closer to every other node by no more than 1, for a total savings of no more than $n - 1$. Consider some other node j which is not a descendant of i (addressed above) and

not its parent (trivially wasteful to link to). Let j_0 be the sibling of i which is an ancestor of j, or root, if there is no such sibling[3]; then the simple path $j_0, j_1, \ldots, j_x = j$ passes through no node linked to i. We know that the savings from linking to j_0 are strictly less than α — now simply note that the savings from linking to j_{y+1} are even less than those from linking to j_y, since i is brought closer by 1 to the nodes in j_{y+1}'s subtree, and moved further by 1 from the nodes in the subtrees of at least 2 other children of j_y (the ones that are the ancestors of neither j nor i, and hence do not have any shorter paths to i). Inductively, the savings from linking to j are strictly less than α as well. Furthermore, since making any one extra link alone is not worth it, making more than one cannot be worthwhile either, since the net savings in the distance component are no greater than the sum of the savings yielded by adding any link alone. Since none of the edges i is paying for can be removed either, $T_{k,d}$ is a Nash equilibrium. $\square$

The total cost of the social optimum (the star, for sufficiently high values of d and n) is $\alpha(n-1) + 2n(n-1) = (d+1)n(n-1)$. By counting distances between leaves alone, which number at least $n(k-1)/k$, we get $C(T_{d,k}) \geq \alpha(n-1) + 2d(\frac{n(k-1)}{k} - 1)^2$. Thus, $\lim_{k,d \to \infty} \rho(T_{d,k}) = \lim_{d \to \infty}((d-1)n(n-1) + 2d(n-1)^2)/((d+1)n(n-1)) = 3$.

5. THE TREE CONJECTURE

Numerous experiments and attempted constructions for $\alpha > 2$ have thus far only yielded Nash equilibria that are trees, with the sole exception of the Petersen graph, which is a transient equilibrium for $1 \leq \alpha \leq 4$. Furthermore, it is somewhat intuitive to suppose that non-tree equilibria are less likely to appear as alpha grows and redundant links become costlier[4]. We thus postulate the following:

CONJECTURE 1 (THE TREE CONJECTURE). *There exists a constant A, such that for all $\alpha > A$, all non-transient Nash equilibria are trees.*

Armed with this conjecture, we can exploit the structure of trees to obtain a constant upper bound on the price of anarchy.

THEOREM 3. *For any tree Nash equilibrium T, $\rho(T) < 5$.*

PROOF. Let n be the number of nodes in T, and, for a node i, let $L(i)$ be the set of nodes which forms the largest connected component of the graph obtained if i and all edges adjacent to it are removed — these components are hereafter referred to as i's subtrees. An arbitrary one is chosen for $L(i)$ if there is more than one subtree of i of the maximum size. We first note that T has at least one center node z — a node such that $|L(z)| \leq n/2$. This may be seen by starting at an arbitrary node z_0, and following a sequence of nodes $z_0, z_1, \ldots$ until a center is reached; z_i is obtained from (non-center) z_{i-1} by following the edge between z_{i-1} and $L(z_{i-1})$. If we consider the subtrees of z_i, all but the one containing z_{i-1} will be strict subsets of $L(z_{i-1})$, while the subtree containing z_{i-1} will be the complement of $L(z_{i-1})$ and, since $|L(z_{i-1})| > n/2$, will contain at most $n/2$ nodes. Hence, $|L(z_i)| \leq \max\{n/2, |L(z_{i-1})| - 1\}$, which clearly guarantees that the sequence will eventually reach a center node.

Let $d \geq 2$ be the depth of T when rooted at z (if $d = 1$, we have a star, for which the theorem holds trivially). That is, some leaf l at depth d chose not to link to z. Since the subtree of z within which l lies contains at most $n/2$ nodes, there are at least $n/2 - 1$ nodes such that paths from l to them go through z. Thus a link from l to z would yield a savings of at least $(d-1)(n/2-1)$ for c_l. Hence, $\alpha \geq (d-1)(n/2-1)$.

The diameter of T is at most $2d$, so $\text{diam}(T) < 2 + 4\alpha/(n-2)$. Then, since $2(n-1)$ ordered pairs of nodes are at distance 1 from each other,

$$-C(T) \leq \alpha(n-1) + \left(2 + \frac{4\alpha}{n-2}\right)(n-2)(n-1) + \tag{2}$$
$$+ 2(n-1) = 5\alpha(n-1) + 2(n-1)^2.$$

Thus, the price of anarchy is

$$\rho(T) \leq \frac{5\alpha(n-1) + 2(n-1)^2}{\alpha(n-1) + 2n(n-1)} < 5$$

. $\square$

From Theorem 1 and Theorem 3, we directly obtain:

COROLLARY 4. *If the Tree Conjecture holds for some value of A, then for all non-transient Nash equilibria G, $\rho(G) = O(\sqrt{A})$.*

It should be noted that the above proof relies crucially on the existence of a "center" node, which seems to be the primary barrier to extending the proof to more general graphs.

Note also that the non-transience constraint in the conjecture and the consequent results is optional, as we have not found even transient non-tree equilibria for $\alpha > 4$. It is included above because it weakens the conjecture and seems to provide a more natural notion of an equilibrium.

6. DISCUSSION AND EXTENSIONS

We would love, of course, to see a proof of the Tree Conjecture, as well as any other constant upper bound for the price of anarchy.

There are some interesting variations on our model that may be worthy of study:

---

[3]Note that this includes the case when i and j share a non-root ancestor other than a parent.

[4]Note that, for a position to be a Nash equilibrium, there must be no single node with *a* different strategy that lowers its costs, no matter how different that strategy is from its current strategy. E.g., it may at first appear that a directed n-cycle for $n \geq 5$ (i.e. a situation where node i links to $i+1$ only, with $n-1$ linking to 0), is a Nash equilibrium. This is not the case, since node 0 would lower its costs if it instead linked to node 2 only.

- In our model, edges paid for by a node can be used by others. In the *directed version* an edge paid by i can only be used in the direction from i to j. The equilibria in this model are much more complex and numerous, although this variant seems to be less applicable to real situations.

- In our model, a node must either pay for an edge in full, or not at all (and hope that the other endpoint will decide to pay for it). In the *fractional model*, a node may pay for a fraction of an edge, and the edge will exist if the sum of the two fractions exceed one (naturally, at equilibrium this sum will be either zero or one). This model may allow for a greater variety of equilibria.

- Stars (even though prevalent in the real Internet) are problematic networks because of congestion. It would be interesting to introduce this consideration to our model; the challenge again is, in doing so, to keep it relatively tractable.

- We assume that all pairs of nodes have the same traffic, weighing distances between all node pairs equally. A more detailed model would use a *traffic matrix* $[t_{ij}] \geq 0$, and the term $d_{G(s)}(i, j)$ of cost function would be multiplied by t_{ij}.

- The previous extension uses n^2 parameters, a little too many. In the *rank-one* special case we assume that $t_{ij} = w_i \cdot w_j$, where w_i are node weights capturing the strength (or importance, or market share) of the nodes. It would be *very* interesting to identify ways in which the weights affect the degrees of the nodes at equilibrium; such an approach may shed light to phenomena of heavy-tailed distributions recently observed [2] (it would be a much more primitive assumption, consistent with experience from other markets, to postulate that the *weights* are so distributed).

- With our colleague Kunal Talwar, we have been considering a network creation game in which node costs reflect *Vickrey payments* — a proposal from [3]. The nature of the equilibria of such a game could explain the modest Vickrey overcharges observed in the real Internet in [3].

- A more realistic model would assume a large but *not infinite* cost to i when j is disconnected from the network. Such an assumption may disallow certain absurd equilibria that are the result of "blackmail by j."

- Finally, it would be more realistic to look at networks that are developed in *stages*, where new nodes arrive at each stage (old nodes can add links), and all stages should be equilibria.

While some of these variants and extensions seem analytically intractable, we believe that research in some of these more sophisticated models may result in increased understanding of the fascinating phenomenon that is the Internet.

7. REFERENCES

[1] A. Czumaj, P. Krysta, and B. Vöcking. Selfish traffic allocation for server farms. In *Proc. of 2002 STOC*, pages 287–296. ACM, May 2002.

[2] M. Faloutsos, P. Faloutsos, and C. Faloutsos. On power-law relationships of the internet topology. In *Proc. of 1999 SIGCOMM*, pages 251–262. ACM, September 1999.

[3] J. Feigenbaum, C. Papadimitriou, R. Sami, and S. Shenker. A bgp-based mechanism for lowest-cost routing. In *Proc. of 2002 PODC*, pages 173–182. ACM, July 2002.

[4] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. *Lecture Notes in Computer Science*, 1563:404–413, 1999.

[5] M. Mavronikolas and P. Spirakis. The price of selfish routing. In *Proc. of 2001 STOC*, pages 510–519. ACM, July 2001.

[6] T. Roughgarden. *Selfish Routing*. PhD thesis, Cornell University, 2002.

[7] T. Roughgarden and E. Tardos. How bad is selfish routing? In *Proc. of 2000 FOCS*, pages 93–102. IEEE, November 2000.