On the Topologies Formed by Selfish Peers

Mocsibroda, Schmid, Wattenhofer

Social Networks Seminar Fall 2007

Oct 16, 2007

Introduction

- Peer-to-peer systems
  - Collaboration of peers to build scalable decentralized systems
  - e.g. BitTorrent for file distribution

Introduction

- Peer-to-peer systems
  - Collaboration of peers to build scalable decentralized systems
  - e.g. BitTorrent for file distribution
- Problem: many peers are *selfish*
  - Minimize own cost

Introduction

- Peer-to-peer systems
  - Collaboration of peers to build scalable decentralized systems
  - e.g. BitTorrent for file distribution
- Problem: many peers are *selfish*
  - Minimize own cost
- Use a *game-theoretic* approach to study the effect of selfish peers on the topology of a P2P system

Introduction

- Peer-to-peer systems
  - Collaboration of peers to build scalable decentralized systems
  - e.g. BitTorrent for file distribution
- Problem: many peers are *selfish*
  - Minimize own cost
- Use a *game-theoretic* approach to study the effect of selfish peers on the topology of a P2P system
  - Price of anarchy: $\Theta(\min(\alpha, n))$
  - Topologies may not converge to a stable state (Nash equilibrium)

The Model

- Set of n peers: $V = \{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$

The Model

- Set of n peers: $V = \{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$
- Distance function $d : V \times V \rightarrow [0, \infty)$
  - Underlying latencies between all pairs of peers
  - Metric space $\mathcal{M} = (V, d)$

The Model

- Set of n peers: $V = \{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$
- Distance function $d : V \times V \to [0, \infty)$
  - Underlying latencies between all pairs of peers
  - Metric space $\mathcal{M} = (V, d)$
- Network Formation Game
  - Strategy space of π_i: $S_i = 2^{V \setminus \{\pi_i\}}$
  - π_i maintains or establishes a link to π_j if $\pi_j \in s_i$, where $s_i \in S_i$
  - $s = (s_0, \ldots, s_{n-1}) \in S_0 \times \cdots \times S_{n-1}$ yields a directed graph $G[s] = \left(V, \bigcup_{i=0}^{n-1} (\{\pi_i\} \times s_i)\right)$ describing the P2P topology

The Model

- Set of n peers: $V = \{\pi_0, \pi_1, \ldots, \pi_{n-1}\}$
- Distance function $d : V \times V \to [0, \infty)$
  - Underlying latencies between all pairs of peers
  - Metric space $\mathcal{M} = (V, d)$
- Network Formation Game
  - Strategy space of π_i: $S_i = 2^{V \setminus \{\pi_i\}}$
  - π_i maintains or establishes a link to π_j if $\pi_j \in s_i$, where $s_i \in S_i$
  - $s = (s_0, \ldots, s_{n-1}) \in S_0 \times \cdots \times S_{n-1}$ yields a directed graph $G[s] = \left(V, \bigcup_{i=0}^{n-1}(\{\pi_i\} \times s_i)\right)$ describing the P2P topology

In plain English, each peer decides whether to establish a direct link to each of its other peers
$\Rightarrow$ a directed graph with peers as nodes and links as edges

The Model

- Each peer has a cost function

$$c_i(s) = \alpha \cdot |s_i| + \sum_{i \neq j} stretch_{G[s]}(\pi_i, \pi_j)$$

where

$$stretch_{G[s]}(\pi_i, \pi_j) = \frac{d_G(\pi, \pi')}{d(\pi, \pi')}$$

The Model

- Each peer has a cost function

$$c_i(s) = \alpha \cdot |s_i| + \sum_{i \neq j} stretch_{G[s]}(\pi_i, \pi_j)$$

where

$$stretch_{G[s]}(\pi_i, \pi_j) = \frac{d_G(\pi, \pi')}{d(\pi, \pi')}$$

  - d is the direct distance, d_G is the sum of underlying latencies along the shortest path on G

The Model

- Each peer has a cost function

$$c_i(s) = \alpha \cdot |s_i| + \sum_{i \neq j} stretch_{G[s]}(\pi_i, \pi_j)$$

where

$$stretch_{G[s]}(\pi_i, \pi_j) = \frac{d_G(\pi, \pi')}{d(\pi, \pi')}$$

  - d is the direct distance, d_G is the sum of underlying latencies along the shortest path on G
  - α is a parameter measuring the relative importance between node degree and stretch in the cost function

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology
  - Maintaining links takes up bandwidth

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology
  - Maintaining links takes up bandwidth
    $\Rightarrow$ Larger $\alpha \cdot |s_i|$ (bad)

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology
  - Maintaining links takes up bandwidth
    $\Rightarrow$ Larger $\alpha \cdot |s_i|$ (bad)
- Disconnected peers communicate through chains of peers $\equiv$ Traversing links on the P2P topology

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology
  - Maintaining links takes up bandwidth $\Rightarrow$ Larger $\alpha \cdot |s_i|$ (bad)
- Disconnected peers communicate through chains of peers $\equiv$ Traversing links on the P2P topology
  - Long chains are slow

Physical Interpretation

- Peers i and j connect with a TCP connection $\equiv$ A direct link on the P2P topology
  - Maintaining links takes up bandwidth
    $\Rightarrow$ Larger $\alpha \cdot |s_i|$ (bad)
- Disconnected peers communicate through chains of peers $\equiv$ Traversing links on the P2P topology
  - Long chains are slow
    $\Rightarrow$ Larger $\alpha \cdot |s_i|$ (bad)

Considerations

- Every peer tries to minimize its cost in the game

Considerations

- Every peer tries to minimize its cost in the game
- Nash equilibrium is the strategy combination s such that for any other combination s' which differ in only peer i's strategy, $c_i(s) \leq c_i(s')$

Considerations

- Every peer tries to minimize its cost in the game
- Nash equilibrium is the strategy combination s such that for any other combination s' which differ in only peer i's strategy, $c_i(s) \leq c_i(s')$
  - No peer has the incentive to deviate (because it cannot gain by deviating)
  - Pure Nash equilibrium

Considerations

- Every peer tries to minimize its cost in the game
- Nash equilibrium is the strategy combination s such that for any other combination s' which differ in only peer i's strategy, $c_i(s) \leq c_i(s')$
    - No peer has the incentive to deviate (because it cannot gain by deviating)
    - Pure Nash equilibrium
- Social cost

$$C(G) = \sum_i c_i = \alpha|E| + \sum_{i \neq j} stretch_G(\pi_i, \pi_j)$$

Considerations

- Every peer tries to minimize its cost in the game
- Nash equilibrium is the strategy combination s such that for any other combination s' which differ in only peer i's strategy, $c_i(s) \leq c_i(s')$
  - No peer has the incentive to deviate (because it cannot gain by deviating)
  - Pure Nash equilibrium
- Social cost

$$C(G) = \sum_i c_i = \alpha|E| + \sum_{i \neq j} stretch_G(\pi_i, \pi_j)$$

- Price of anarchy =

$$\frac{\text{social cost of worst NE}}{\text{social cost of optimal topology}}$$

Upper Bound

Theorem 4.1 (Upper bound)

For any metric space $\mathcal{M}$, the Price of Anarchy is $O(\min(\alpha, n))$

Proof

Upper Bound

Theorem 4.1 (Upper bound)

For any metric space $\mathcal{M}$, the Price of Anarchy is O(min(α, n))

Proof

- $1 \leq stretch(\pi, \pi') \leq \alpha + 1$ (if $> \alpha + 1$, π can establish direction connection to π', reducing cost from $> \alpha + 1$ to $\alpha + 1$)

Upper Bound

Theorem 4.1 (Upper bound)

For any metric space $\mathcal{M}$, the Price of Anarchy is $O(\min(\alpha, n))$

Proof

- $1 \leq stretch(\pi, \pi') \leq \alpha + 1$ (if $> \alpha + 1$, π can establish direction connection to π', reducing cost from $> \alpha + 1$ to $\alpha + 1$)
- Max social cost $= \alpha \cdot n(n-1) + n^2(\alpha + 1) = O(\alpha n^2)$
  - $n(n-1)$ is the maximum number of edges in G
  - n^2 is the number of (i, j) pairs in G

Upper Bound

Theorem 4.1 (Upper bound)

For any metric space $\mathcal{M}$, the Price of Anarchy is $O(\min(\alpha, n))$

Proof

- $1 \leq \mathit{stretch}(\pi, \pi') \leq \alpha + 1$ (if $> \alpha + 1$, π can establish direction connection to π', reducing cost from $> \alpha + 1$ to $\alpha + 1$)
- Max social cost $= \alpha \cdot n(n-1) + n^2(\alpha + 1) = O(\alpha n^2)$
  - $n(n-1)$ is the maximum number of edges in G
  - n^2 is the number of (i, j) pairs in G
- Min social cost $= \alpha n + n^2 \cdot 1 = \Omega(\alpha n + n^2)$
  - n is the minimum number of edges (G must be connected)

Upper Bound

Theorem 4.1 (Upper bound)

For any metric space $\mathcal{M}$, the Price of Anarchy is $O(\min(\alpha, n))$

Proof

- $1 \leq \mathit{stretch}(\pi, \pi') \leq \alpha + 1$ (if $> \alpha + 1$, π can establish direction connection to π', reducing cost from $> \alpha + 1$ to $\alpha + 1$)
- Max social cost $= \alpha \cdot n(n-1) + n^2(\alpha + 1) = O(\alpha n^2)$
  - $n(n-1)$ is the maximum number of edges in G
  - n^2 is the number of (i, j) pairs in G
- Min social cost $= \alpha n + n^2 \cdot 1 = \Omega(\alpha n + n^2)$
  - n is the minimum number of edges (G must be connected)
- Price of anarchy $= \frac{O(\alpha n^2)}{\Omega(\alpha n + n^2)} = O(\min(\alpha, n))$ $\square$

Lower Bound

- Lower bound not applicable to all metric spaces

Lower Bound

- Lower bound not applicable to all metric spaces
- Here we consider $\mathcal{M}$ described by n points on a real line (1D Euclidean space)

Lower Bound

- Lower bound not applicable to all metric spaces
- Here we consider $\mathcal{M}$ described by *n* points on a real line (1D Euclidean space)

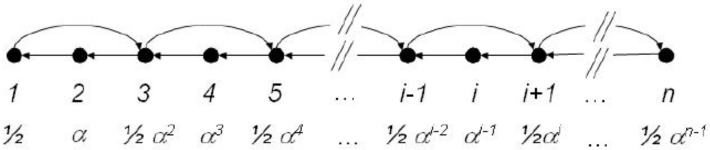


Figure 1: Example topology G where the Price of Anarchy is $\Theta(\min(\alpha, n))$ for $3.4 \le \alpha$.

Lower Bound

Lemma 4.2

The topology G shown in Figure 1 forms a Nash equilibrium for $\alpha \geq 3.4$

Lemma 4.3

The social cost $C(G)$ of the topology G shown in Figure 1 is $C(G) \in \Theta(\alpha n^2)$

Theorem 4.4

The Price of Anarchy of the peer topology G shown in Figure 1 is $\Theta(\min(\alpha, n))$

Main Theorem

Theorem 5.1

Regardless of the magnitude of α, there are metric spaces $\mathcal{M}$, for which there exists no pure Nash equilibrium, i.e. certain P2P networks cannot converge to a stable state. This is the case even if $\mathcal{M}$ is a 2-dimensional Euclidean space.

Illustration

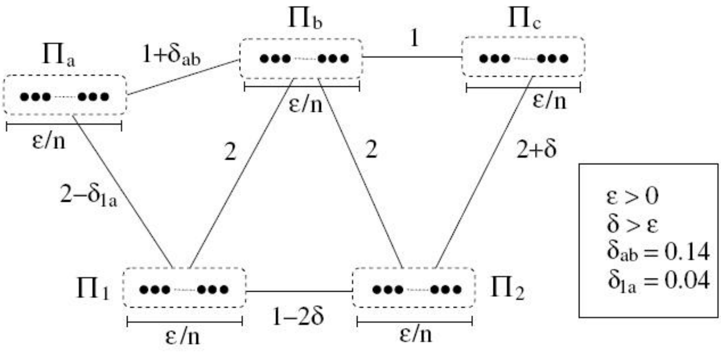


Figure 2: Instance I_k has no pure Nash equilibrium when $\alpha = 0.6k$, where $k = n/5$. The number of peers in each cluster is k.

Complexity

Theorem 6.1

Regardless of the magnitude of α, determining whether a given P2P network represented by a metric space $\mathcal{M}$ has a pure Nash equilibrium (and can therefore stabilize) is NP-hard.

Conclusion

- Comparison with Fabrikant et al's paper
  - Directed instead of undirected links
  - Consider underlying latencies instead of hop count
  - Proved pure Nash equilibria may not exist

Conclusion

- Comparison with Fabrikant et al's paper
  - Directed instead of undirected links
  - Consider underlying latencies instead of hop count
  - Proved pure Nash equilibria may not exist
- Game-theoretic model of a P2P system

Conclusion

- Comparison with Fabrikant et al's paper
  - Directed instead of undirected links
  - Consider underlying latencies instead of hop count
  - Proved pure Nash equilibria may not exist
- Game-theoretic model of a P2P system
- Bounds of Price of Anarchy

Conclusion

- Comparison with Fabrikant et al's paper
  - Directed instead of undirected links
  - Consider underlying latencies instead of hop count
  - Proved pure Nash equilibria may not exist
- Game-theoretic model of a P2P system
- Bounds of Price of Anarchy
- (Non-)Existence of Nash equilibria

Implications of Selfish Neighbor Selection in Overlay Networks

N. Laoutaris, G. Smaragdakis, A. Bestavros and J. Byers

Social Networks Seminar

October 16th, 2007

Motivation

- Neighbor Selection: Key problem.

Motivation

- Neighbor Selection: Key problem.
- Previous Work:
  - undirected edges
  - hop-count based cost
  - unbounded degree
  - PRACTICAL INTEREST?

Motivation

- Neighbor Selection: Key problem.
- Previous Work:
  - undirected edges
  - hop-count based cost
  - unbounded degree
  - PRACTICAL INTEREST?
- Players do not play in a socially optimal way: selfish

Scope and Contributions

- More practical model: for P2P and overlay routing application
  - Bounded Degree
  - Directed Edges
  - Non-uniform Preference Vectors
  - Weighted Edges

Scope and Contributions

- More practical model: for P2P and overlay routing application
  - Bounded Degree
  - Directed Edges
  - Non-uniform Preference Vectors
  - Weighted Edges
- Wiring Strategy: "Best Response"

Scope and Contributions

- More practical model: for P2P and overlay routing application
  - Bounded Degree
  - Directed Edges
  - Non-uniform Preference Vectors
  - Weighted Edges
- Wiring Strategy: "Best Response"
- Experimental Results

Definitions

- Set of nodes: $V = \{v_1, v_2, \dots, v_n\}$
- Edge $e = \{v_i, v_j\}$: *Directed and Weighted*
- Preference vector: $p_i = \{p_{i1}, p_{i2}, \dots, p_{ii-1}, p_{ii+1}, p_{in}\}$
- Wiring of v_i: $s_i = \{v_{i_1}, v_{i_2}, \dots, v_{i_{k_i}}\}$
- Global Wiring: $S = \{s_1, s_2, \dots, s_n\}$
- Cost between v_i, v_j over S: $d_S(v_i, v_j)$
- Cost of v_i under S: $C_i(S) = \sum_{j=1, j \neq i}^{n} p_{ij} \times d_S(v_i, v_j)$

Definitions

- The Selfish Neighbor Selection (SNS) Game: $\langle V, \{S_i\}, \{C_i\}\rangle$
- Residual wiring: $S_{-i} = S - \{s_i\}$

Definitions

- The Selfish Neighbor Selection (SNS) Game: $\langle V, \{S_i\}, \{C_i\} \rangle$
- Residual wiring: $S_{-i} = S - \{s_i\}$
- Best Response: A wiring $s_i \in S_i \ s.t. \ C_i(S_{-i} + \{s_i\}) \leq C_i(S_{-i} + \{s_{i'}\}), \forall s_i' \neq s_i$

Definitions

- The Selfish Neighbor Selection (SNS) Game: $\langle V, \{S_i\}, \{C_i\} \rangle$
- Residual wiring: $S_{-i} = S - \{s_i\}$
- Best Response: A wiring $s_i \in S_i \; s.t. \; C_i(S_{-i} + \{s_i\}) \leq C_i(S_{-i} + \{s_{i'}\}), \forall s_i' \neq s_i$
- Stable Wiring: A global wiring S is **stable** *iff* all individual wirings are **best responses**.

Definitions

- The Selfish Neighbor Selection (SNS) Game: $\langle V, \{S_i\}, \{C_i\} \rangle$
- Residual wiring: $S_{-i} = S - \{s_i\}$
- Best Response: A wiring $s_i \in S_i \; s.t. \; C_i(S_{-i} + \{s_i\}) \leq C_i(S_{-i} + \{s_{i'}\}), \forall s_i' \neq s_i$
- Stable Wiring: A global wiring S is **stable** *iff* all individual wirings are **best responses**.

  (Nash Equilibrium of the SNS game)

Modeling Wirings

- For a node v_i
  - $Y_l \in \{0,1\}^{n-1}$: if v_i wires to v_l
  - $X_{lj} \in \{0,1\}^{n-1}$: if v_i has v_l as a first-hop neighbor to v_j

Modeling Wirings

- For a node v_i
  - $Y_l \in \{0,1\}^{n-1}$: if v_i wires to v_l
  - $X_{lj} \in \{0,1\}^{n-1}$: if v_i has v_l as a first-hop neighbor to v_j
- BR for v_i under S_{-i}, an *Integer Linear Program*
  - Minimize:

$$C_i(S_{-i}, X) = \sum_{j=1, j\neq i}^{n} p_{ij} \sum_{l=1, l\neq i}^{n} X_{lj} \times (d_{il} + d_{S_{-i}}(v_l, v_j))$$

  - Subject to:

$$\sum_{l=1, l\neq i}^{n} X_{lj} = 1, \forall j \neq i \textit{ and } \sum_{l=1, l\neq i}^{n} Y_l = k_i \textit{ and } X_{ij} \leq Y_l, \forall l, j \neq i$$

Modeling Wirings

- For a node v_i
  - $Y_l \in \{0,1\}^{n-1}$: if v_i wires to v_l
  - $X_{lj} \in \{0,1\}^{n-1}$: if v_i has v_l as a first-hop neighbor to v_j
- BR for v_i under S_{-i}, an *Integer Linear Program*
  - Minimize:
$$C_i(S_{-i}, X) = \sum_{j=1, j\neq i}^{n} p_{ij} \sum_{l=1, l\neq i}^{n} X_{lj} \times (d_{il} + d_{S_{-i}}(v_l, v_j))$$
  - Subject to:
$$\sum_{l=1, l\neq i}^{n} X_{lj} = 1, \forall j \neq i \; and \; \sum_{l=1, l\neq i}^{n} Y_l = k_i \; and \; X_{ij} \leq Y_l, \forall l, j \neq i$$
- NP-Hard

Modeling Wirings

- When all the wires have the same unitary weight...

Modeling Wirings

- When all the wires have the same unitary weight...
- Find a node's BR $\Rightarrow$ Solving a k-median problem

Modeling Wirings

- When all the wires have the same unitary weight...
- Find a node's BR $\Rightarrow$ Solving a k-median problem
- *Definition*: Asymmetric k-median

Modeling Wirings

- When all the wires have the same unitary weight...
- Find a node's BR $\Rightarrow$ Solving a k-median problem
- *Definition*: Asymmetric k-median
  - Select up to k nodes to act as median so as to minimize..

$$C(V', k, w) = \sum_{\forall v_j \in V'} w_j \times d_{S'}(v_j, m(v_j))$$

$m(v_j)$ is the median that is closest to v_j

Modeling Wirings

- When all the wires have the same unitary weight...
- Find a node's BR $\Rightarrow$ Solving a k-median problem
- *Definition*: Asymmetric k-median
  - Select up to k nodes to act as median so as to minimize..

$$C(V', k, w) = \sum_{\forall v_j \in V'} w_j \times d_{S'}(v_j, m(v_j))$$

$m(v_j)$ is the median that is closest to v_j

- NP-Hard

To Find Stable Wirings

- **Iterative Best Response**

  At the m-th iteration, for all nodes, do

  1. v_i computes its best response $s_i^{(m)}$ to $S_{-i}^{(m,i-1)}$
  2. $S^{(m,i)} = S_{-i}^{(m,i-1)} + \{s_i^{(m)}\}$
  3. Do this until iteration $\mathcal{M}$: $s_i^{(\mathcal{M})} = s_i^{(\mathcal{M}-1)}$, $\forall v_i \in V$

Characterization of Stable Wirings

- Two key scaling parameters:

Characterization of Stable Wirings

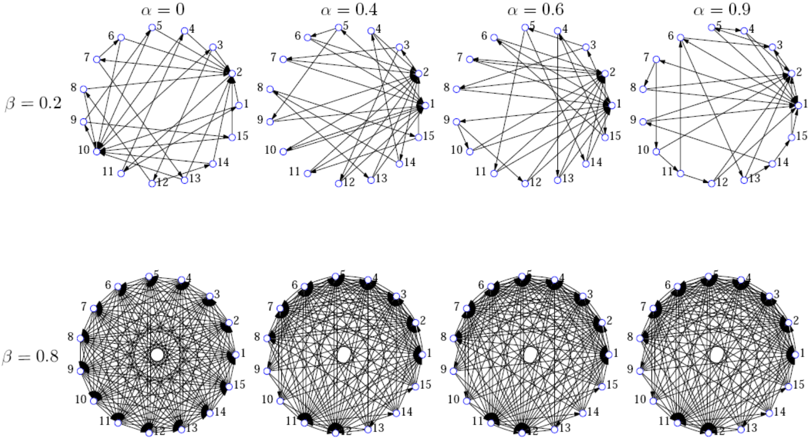
- Two key scaling parameters:
  - $\alpha \in [0, 1]$: non-uniformity (skewness) in popularity (Zipf with parameter α)
    - $q_i = \Lambda / i^{\alpha}$, where $\Lambda = (\sum_{i'=1}^{n} \frac{1}{i'^{\alpha}})^{-1}$

Characterization of Stable Wirings

- Two key scaling parameters:
  - $\alpha \in [0, 1]$: non-uniformity (skewness) in popularity (Zipf with parameter α)
    - $q_i = \Lambda / i^{\alpha}$, where $\Lambda = (\sum_{i'=1}^{n} \frac{1}{i'^{\alpha}})^{-1}$
  - $\beta \in [0, 1]$: link density
    - Out-degree of each node: $k = \lceil n^{\beta} \rceil$

Characterization of Stable Wirings

Note: "Hubs"



Performance of Newcomers: the cost

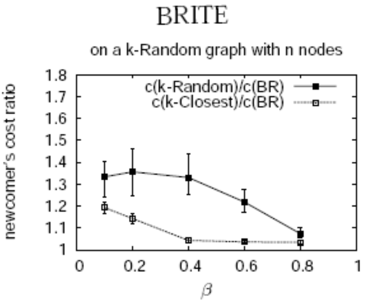
- Three kinds of Wiring Strategies
  - Best Response: selfish
  - k-Closest: selfish
  - k-Random: not very selfish

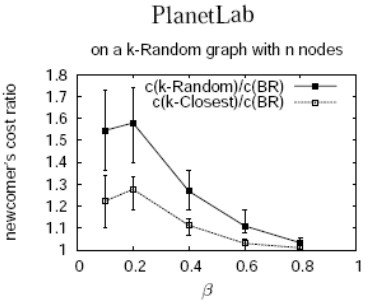
Performance of Newcomers: the cost

- Three kinds of Wiring Strategies
  - Best Response: selfish
  - k-Closest: selfish
  - k-Random: not very selfish
- Newcomers using different strategies entering graphs constructed by different strategies.

Performance of Newcomers

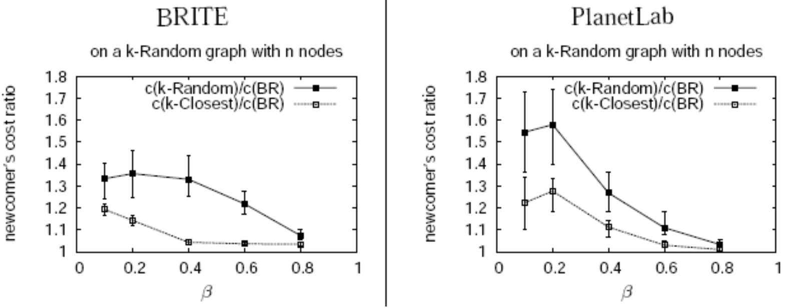
Newcomers to a *k*-random graph





Performance of Newcomers

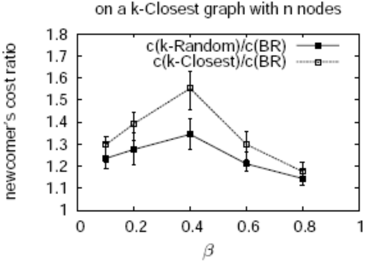
Newcomers to a k-random graph

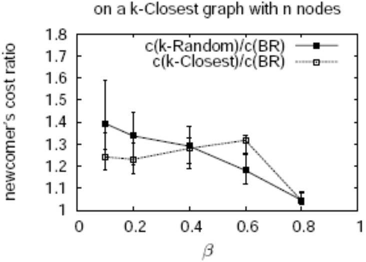


It pays to "cheat".

Performance of Newcomers

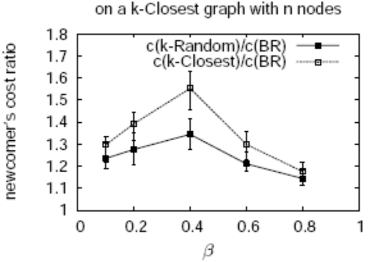
Newcomers to a *k*-closest graph

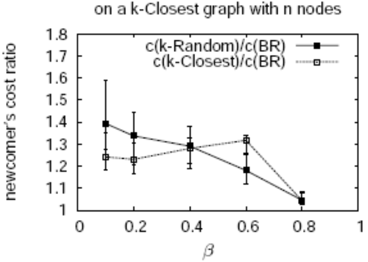




Performance of Newcomers

Newcomers to a *k*-closest graph

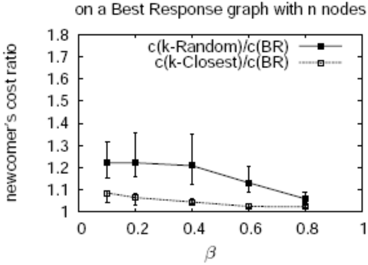


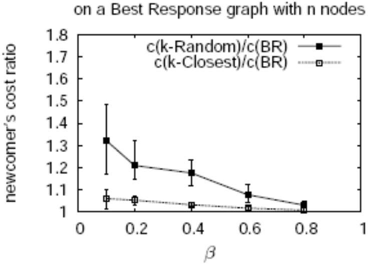


It does NOT pay to cheat, it may even hurt!

Performance of Newcomers

Newcomers to a BR graph





Knowledge Gained

- For prior arrivals: it is good to be selfish.
- For newcomers: should be careful.

Conclusion on Papers **A** and **B**

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game
- **B**: more practical, giving how to wire

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game
- **B**: more practical, giving how to wire
- Common points in topology: directed, weighted edges

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game
- **B**: more practical, giving how to wire
- Common points in topology: directed, weighted edges
- Difference: **A**: unbounded degree **B**: bounded degree

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game
- **B**: more practical, giving how to wire
- Common points in topology: directed, weighted edges
- Difference: **A**: unbounded degree **B**: bounded degree
  - **A**: Trade-off between degree and distance

Conclusion on Papers **A** and **B**

- **A**: theoretical analysis of the properties of the game
- **B**: more practical, giving how to wire
- Common points in topology: directed, weighted edges
- Difference: **A**: unbounded degree **B**: bounded degree
  - **A**: Trade-off between degree and distance
  - **B**: The optimization under certain restrictions

Conclusion on Papers **A** and **B**

- The restrictions make differences

Conclusion on Papers **A** and **B**

- The restrictions make differences
  - **A**: Nash Equilibrium may not exist.

Conclusion on Papers **A** and **B**

- The restrictions make differences
  - **A**: Nash Equilibrium may not exist.
  - **B**: (It seems) we can always find an N.E.

Conclusion on Papers **A** and **B**

- The restrictions make differences
  - **A**: Nash Equilibrium may not exist.
  - **B**: (It seems) we can always find an N.E.
  - **A**: Selfishness can lead to very bad consequences.

Conclusion on Papers **A** and **B**

- The restrictions make differences
  - **A**: Nash Equilibrium may not exist.
  - **B**: (It seems) we can always find an N.E.
  - **A**: Selfishness can lead to very bad consequences.
  - **B**: Sometimes it is good to be selfish.

Conclusion on Papers **A** and **B**

- Extension

Conclusion on Papers **A** and **B**

- Extension
  - The joining (or leaving) of players.

Conclusion on Papers **A** and **B**

- Extension
  - The joining (or leaving) of players.
  - Using different strategies in a single network.