BUDAPEST UNIVERSITY OF TECHNOLOGY AND ECONOMICS
FACULTY OF ELECTRICAL ENGINEERING AND INFORMATICS
DOCTORAL SCHOOL OF COMPUTER SCIENCE

THE NETWORK FORMING EFFECTS OF ROUTING POLICY

Dávid Szabó
M.Sc. in Computer Engineering

Summary of the Ph.D. Dissertation

Supervised by

Dr. András Gulyás
Assistant professor
High Speed Networks Laboratory (HSN Lab)
Dept. of Telecommunications and Media Informatics
Budapest University of Technology and Economics

Budapest, Hungary
2016

1 Introduction

Knowing more about the topology formation of complex networks – like social networks, biological networks or the Internet – has great importance. Such knowledge can help us to gain insight into the decision making process of individuals in the network and to understand the principles that govern the emerging topological features. However, despite the large existing literature from multidisciplinary research areas, our knowledge is still limited and incomplete. In general, we have valuable information about network dynamics from multiple aspects, but still the *topological footprints inflicted by routing policies is an unacquainted aspect*, though, this is one of the most determining factors on the resulting network.

To make this statement clear just consider that in every network where there is communication the preferred routes for the information flows emerge somehow. The preferences that drive this process can be described by a set of rules what we collectively call *routing policy.* The routing policy naturally dictates that only those connections are being created and kept that are remunerative for the efficient communication. Fig. 1 shows a simple example to illustrate this idea.

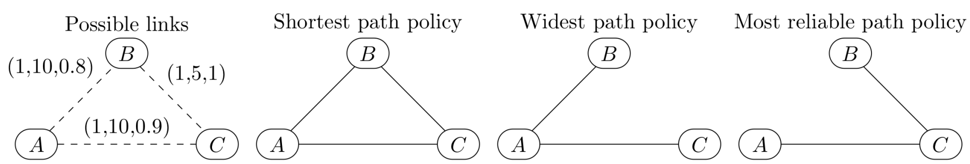


Figure 1. Policy drives topology: We have three nodes (A, B, C) among which we decide to build a network. The characteristics of the possible edges (A, B), (A, C), (B, C) are defined by a triplet: latency, bandwidth and reliability (l, b, r) (rightmost). Now we have to decide what topology should be built. Clearly, if the routing policy is the shortest path policy, then all the links have to be used. If the policy is the widest path, meaning that the highest bandwidth preferred between the nodes, then the third topology from the left is the appropriate choice. Finally, if the policy is the most reliable path policy then the corresponding network will be the leftmost topology.

Based on this influence between the used routing policy and the network topology, we can say that understanding the topological footprint of a routing policy can give valuable insight into the topological properties of a real network.

2 Research Objectives

The objective of my dissertation is to give a comprehensive analysis about the network formation effects of different routing policies related to complex networks. To do this I discuss two fundamental routing policies: first I focus on the *Border Gateway Protocol* that encompasses the inter-domain routing policies on the Internet autonomous system level topology, and within that the two most important rules called *valley-free* (VF) and *highest local preference* (HLP) rule, then I broaden the context from the Internet specific routing policies to investigate *greedy navigation* that is the most accepted routing policy of general complex networks.

In particular, in the first part I define models in an incremental way for the VF and the HLP rules and identify the emerging topologies that can be understood as a direct consequence of these policies. I characterize the topological artifacts of these topologies and try to give reasonable predictions about the Internet's AS level topology based on the models.

In the second part, I design a topology generator, based on the previous results. I show that, although the model is very simple, still the generated topologies reflect the features of the AS level topology along several metrics. I also provide a comprehensive comparison with existing topology generators.

In the third part, I create a model for greedy navigation that describes the communication more realistically than previous models, namely, instead of *shortest path* nodes use *greedy path* forwarding. I give a proof that in the Euclidean space the emergence of topologies on which we understand the principles of greedy navigation and which inspire many algorithms cannot be justified through the interaction of rational, selfish players. I also present a brief outlook on how the situation changes if the hyperbolic space is used, instead of Euclidean.

I believe that my results may contribute to extending our knowledge about the effects of routing policies on the topology formation process of complex networks, since they give complementary insight to the existing results.

3 Methodology

Throughout the dissertation I use a game theoretical approach as it provides a suitable toolkit for analyzing the network-creation process from the aspect of individuals. I define models (games) in which the policy rules exist in their pure and realistic form, carefully separated from other factors that could confuse the analytic inference.

Generally a game can be described by three components: *players*, *strategies* and *payoffs*. Formally, such a game consists of a set $\mathcal{P}$ of players (intelligent rational decision-makers) with cardinality N. Each player u has its own set of possible strategies S_u describing the possibilities how u can act. During the game each player u selects a strategy $s_u \in S_u$ and each state of the game is represented by the strategies of players $s = (s_1, ..., s_N)$. The network creation process is considered finished if the game reaches a Nash equilibrium (NE) state:

Definition 1. *The game is in the state of NE if no player has anything to gain by unilaterally changing its strategy. In other words, a strategy $s \in S$ constitutes a Nash equilibrium if for all players u and for each alternate strategy $s'_u \in S_u : I_u(s_u, s_{-u}) \geq I_u(s'_u, s_{-u})$, where s_u denotes the strategy played by player u, s_{-u} denotes the strategies played by all other players and I_u denotes the payoff of player u.*

In the models the network is represented with a *simple, undirected graph* $G(V, E)$ with V being the set of nodes and E the set of links and all nodes are considered as rational, selfish players whose intention is forming a network along their own interest. During the investigation I first perform a *theoretical analysis*, that is followed by *simulations* and *measurements* (when applicable, e.g. in the case of the Internet). The simulations are implemented in C++/igraph[1] and the results are conducted through various BASH[2] scripts. During the comparison with the real AS level network I use topological information from CAIDA[3]. Since the generated topologies are represented in Geographic Markup Language (GML) format, I converted CAIDA's data to this format.

[1] igraph is library providing a collection of efficient implementations of common data structures and algorithms for network analysis.

[2] Bourne Again SHell is a command-line interface for interacting with Unix/Linux based operating systems.

[3] The Center for Applied Internet Data Analysis (CAIDA) is a collaborative undertaking among organizations in the commercial, government, and research sectors aimed at promoting greater cooperation in the engineering and maintenance of a robust, scalable global Internet infrastructure.

4 New Results

4.1 Consequences of the Border Gateway Protocol

In the following, I consider ASs as rational but selfish players whose incentive is to communicate with each other using the *valley-free* (VF) and the *highest local preference* (HLP) rules for routing policies.

Thesis 1. *I have created a game theoretical model that can address the consequences of the most fundamental BGP policy, called valley-free routing. I have identified a graph that is a natural consequence of valley-free routing and which is omnipresent in the Internet AS level topology as a subgraph. I have extended the model by adding the second BGP rule, called highest local preference, to the analysis and I have disclosed and characterized a further refined version of the graph acquired from the first model that I call the "Spiderweb graph".*

4.1.1 The Valley-Free Game

The VF rule is the most fundamental part of the BGP policy routing, since any valid path between ASs has to be a VF path as well. Accordingly, my first goal is to define and analyze a game to understand how this rule affects the topology.

The policy dictates that AS A can use a link to a neighboring AS B to forward the traffic if and only if either the incoming traffic is from a customer or B is a customer of A [15]. In other words, valley-free compliant paths comprise arbitrary (may be zero) number of customer-provider links, zero or one peer link and again arbitrary provider-customer links strictly in this order (Fig. 2). The valley-free policy is a typical example on how important an economic, i.e. a non-technical, policy constraint could be. In the following, I define and analyze a game to understand how this rule affects the topology.

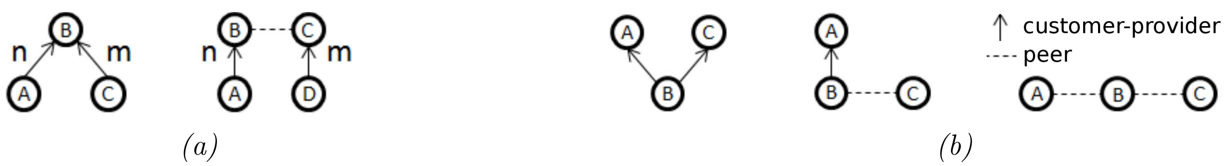


Figure 2. Illustration of path types that (a) satisfy and (b) violate the VF policy. A valid path contains n customer-provider, at most 1 peer and m provider-customer link strictly in this order, where $n, m \in \mathbb{N}$. All the other types are invalid paths.

Thesis 1.1. *I have created a game theoretical model in which the nodes are being incentivized by valley-free routing and I have given a special graph that constitutes the Nash Equilibrium of the game. This special graph comprises a clique as a subgraph – which size depends on the cost ratio of the peering and customer-provider links – and trees rooted at some subset of the clique.*

Players and routing – Let $\mathcal{P}$ be the set of players (identified as network nodes) with cardinality N and the players has incentives to communicate with each other but only via VF paths.

Strategies and topology – A strategy for player $u \in \mathcal{P}$ is to create a set of undirected edges to other players in the network. The created edges can be of types customer-provider (p) and peer (r) edges in accordance with the relationships of the VF routing. The r edges are paid at both sides, however p edges are paid by the customer. Thus the complete strategy space of player u

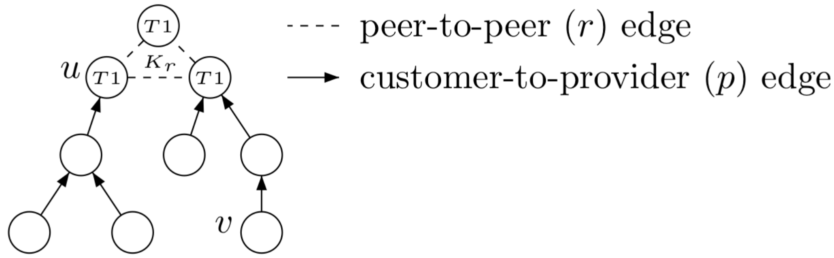


Figure 3. Example for a VFF topology. In such a topology there could be two type of nodes, T1 and none T1. T1s are connected by r edges, which are counted on both sides in the cost function, however p edges are paid only by the customer, who requested it. The flow of cash is visualized by arcs. According to this there are two possible cost functions: (i) $C_u = \varphi_r u_r = \varphi_r(|V(K_r)| - 1)$ *and (ii)* $C_v = \varphi_p$.

is $S_u = 3^{\mathcal{P}\setminus\{u\}}$, where the number 3 covers the third choice of node u, which is creating no edge. Let s be a strategy vector containing the strategies of all players hereby representing the current state of the game: $s = (s_0, s_1, ..., s_{N-1}) \in (S_0, S_1, ..., S_{N-1})$. Then the graph $G(s) = \bigcup_{i=0}^{N-1}(i \times s_i)$ represents the topology between the players.

Payoff – The goal of the players is to minimize their costs. The cost of player u is defined as:

$$C_u(s) = \underbrace{\sum_{v \neq u} d_{G(s)}(u, v)}_{\text{communication cost}} + \underbrace{\varphi_p u_p + \varphi_r u_r}_{\text{link cost}}, \quad u, v \in \mathcal{P}, \tag{1}$$

where φ_x is the cost an edge of type $x \in \{r, p\}$, u_x is the number edges of type x and $d_{G(s)}(u, v)$ is the communication cost between u and v over $G(s)$ given by $d_{G(s)}(u, v) = 0$, if a VF path exists between u and v, otherwise ∞. In what follows I identify the Nash equilibrium of the game in different settings of the parameters.

Definition 2 (Valley-Free footprint (VFF))**.** *A graph is a valley-free footprint if it consists of (i) a clique* K_r *comprising peer (r) links only, and (ii) trees rooted at some subset of* $V(K_r)$ *having customer-provider links (p) only, such that for all provider-customer connections the provider is always closer to their respective root than the customer (see Fig. 3).*

Theorem 1. *A VFF is a Nash equilibrium if and only if* $\left\lceil \frac{\varphi_p}{\varphi_r} \right\rceil \leq |V(K_r)| \leq \left\lfloor \frac{\varphi_p}{\varphi_r} \right\rfloor + 1.$

According to Theorem 1 in nontrivial cases this very simple game exhibits a significant level of structural resemblance to the Internet AS level topology. On the AS level tier-1 ASs (T1s) are in a clique that have peering agreements with each other, this is the top of the hierarchy. The rest of the nodes are customers of T1s either in a direct or in an indirect way. These topological features are clearly reflected by the results and now they can be understood as a clear consequence of the VF policy. Theorem 1 also gives a rough estimation on the number of T1 nodes as the function of edge costs.

4.1.2 The Highest Local Preference Game

The second rule of the Best Path Selection Algorithm is the other very important economically motivated policy, which is the *highest local preference* policy. It is applied on top of valley-free routes meaning that an AS can pick one from the available valley-free routes according to its local interest. Meanwhile these local interests can exhibit a high variety the minimalistic rule, that customer and peer paths are favored over provider paths, is contained in basically every local preference setting within the ASs [11]. Fig. 4 shows a simple illustration in which AS C can reach AS G through multiple paths (C–B–A–D–G, C–D–G and C–E–F–G), however, economically, the favorable order is C–E–F–G (as E is a customer of C), C–D–G (as D is a peer partner of C) and C–B–A–D–G (as B is a provider of C). This is in line with the nature of these routes as customer and peer paths are completely free unlike provider paths in which the provider has to be compensated in some way for the carried transit traffic.

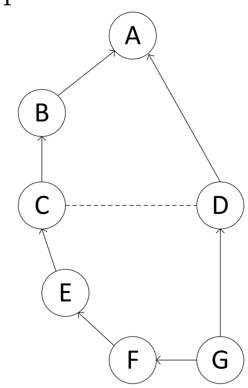


Figure 4. Illustration of paths prioritization according to the highest local preference rule.

s_{vu} \ s_{uv}	0	p	r
0	0	$\overrightarrow{uv}$	0
p	$\overrightarrow{vu}$	$\overline{uv}$	$\overline{uv}$
r	0	$\overline{uv}$	$\overline{uv}$

Figure 5. The created edge according to the strategies of players u and v.

Thesis 1.2. *I have extended the model by adding the second BGP rule called highest local preference to the analysis and I have disclosed and characterized a further refined version of the graph that I call the "Spiderweb graph".*

Players and routing – Let $\mathcal{P}$ be the set of players (identified as the ASs) with cardinality N. Recalling the rule of HLP policy a player always picks from the available VF paths according to its local interest, which is a preference ordering based on the first edge of the path. In this game I use the notations p (or $\overrightarrow{uv}$) and r (or $\overline{uv}$) to denote *customer-provider* and *peer* edges, respectively. This addition is important in order to keep the analysis clear and simple, as in several times referring to edges with their endpoints - instead only their type - is preferable.

Strategies and topology – A strategy for player $u \in \mathcal{P}$ is a vector of the preferred edges to other players in the AS network; i.e. the strategy space is the set $S_u = \{(s_{uv})_{v \in \mathcal{P}\setminus\{u\}} : s_{uv} \in \{0, p, r\}\}$ where $|S_u| = 3^{N-1}$. Easily, player u seeks to contact player v if $s_{uv} \in \{p, r\}$, otherwise $s_{uv} = 0$. Players announce their strategies simultaneously. Any state of the game is represented by an undirected graph $G(s) = (\mathcal{P}, E(s))$ generated by the strategies of the nodes, where $E(s)$ is given by $E(s) = \{\overrightarrow{uv} | s_{uv} = p \wedge s_{vu} = 0\} \cup \{\overline{uv} | s_{uv} \in \{r, p\} \wedge s_{vu} \in \{r, p\}\}$. This settlement of the edges reflects the rational behavior of the ASs as they prefer to create peer edges over customer-provider edges and the instantiation of peer edges requires a *bilateral* agreement between the corresponding players while customer-provider edges can be created *unilaterally*. These can be summarized in Fig. 5.

Payoff – The goal of the players is to minimize their costs, which for a given player u is defined as:

$$C_u(s) = \underbrace{\frac{1}{N}\sum_{v \neq u} d_{G(s)}(u, v)}_{\text{communication cost}} + \underbrace{\varphi_p u_p + \varphi_r u_r}_{\text{link cost}}, \quad v \in \mathcal{P} \tag{2}$$

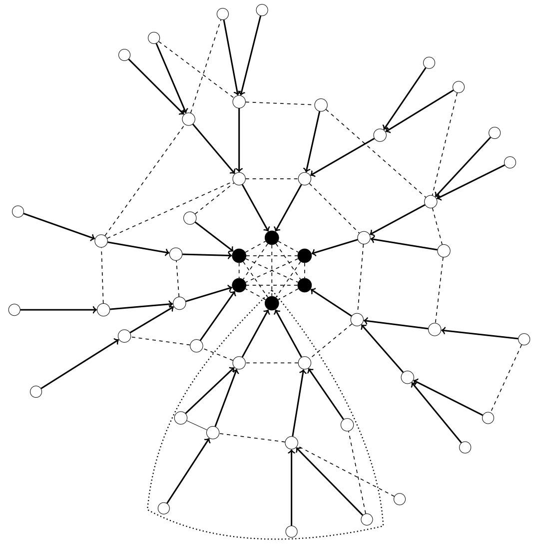


Figure 6. An example of the spiderweb graph, the dashed and directed edges are the peer and customer-provider edges, respectively and the black nodes are the ASs of the clique K, i.e. the tier-1 ASs. The dotted triangle indicates the customer cone of a tier-1 AS.

where

$$d_{G(s)}(u,v) = \begin{cases} 0 & \text{if there exists a VF path whose first edge is } \textit{peer} \text{ or } \textit{provider-customer} \\ 1 & \text{if there exists at least one VF path and the first edge of all of them} \\ & \text{is } \textit{customer-provider} \\ \infty & \text{if a VF path does not exist} \end{cases} \quad (3)$$

represents the price of communication between u and v over $G(s)$ in compliance with the VF and HLP policies, φ_p and φ_r are fix maintenance costs of the provider and peer edges, while u_p and u_r refer to the number of the p and r edges of u, respectively. Note that provider-customer edges are considered to be financed unilaterally by the customer.

In order to find topologies that are more relevant to a realistic network game I used the following more natural and slightly tailored equilibrium definition for this case:

Definition 3 (Pairwise Stable Nash Equilibrium (PSNE) [16])**.** *We say $G(s)$ constitutes a pairwise stable Nash equilibrium if (a) it is a Nash equilibrium, (b) $\forall uv \in E(G(s)) : C_u(s) \le C_u(s') \wedge C_v(s) \le C_v(s')$, where s' differs from s only in deleting one uv edge from $G(s)$, (c) $\forall uv \notin E(G(s)) : C_u(s) \le C_u(s') \vee C_v(s) \le C_v(s')$, where s' differs from s only in adding uv edge to $G(s)$ and (d) contains no provider loops (cycle of p edges)[4].*

Now I am interested in the equilibrium topologies of the game as these topologies will reflect the consequences of the VF and the HLP rules. For the claims the following definition is needed.

Definition 4 (Spiderweb graph (Fig. 6))**.** *A graph is a Spiderweb graph if it consists of:*

1. *a clique K_r (representing the tier-1 ASs) comprising peer edges only*
2. *trees rooted at some subset of $V(K_r)$ that have customer-provider edges such that the provider in the connection is always closer to the root than the customer*
3. *additional peer edges, such that $\forall \overline{uv}, \overline{uw} \in G(s) : t(v) \cap t(w) = \emptyset$, where $t(x)$ is the set of nodes in the subtree (i.e. the customer cone) of node x, including x itself.*

[4]This requirement is fully in line with the Gao-Rexford conditions [10] ensuring BGP stability.

The first claim characterizes all meaningful states (i.e. where all the ASs can communicate with each other) of the above game (and thus the AS topology) by identifying a graph that is omnipresent in the Internet as a subgraph.

Theorem 2. *Every meaningful outcome of the game, i.e.,* $\sum C_u \neq \infty$ *contains the Spiderweb graph as a spanning subgraph and every pairwise stable equilibrium (PSNE) of the game is the Spiderweb graph itself.*

The following theorem gives a high-level insight into the placement of the peer edges.

Theorem 3. *If* $G(s)$ *constitutes a pairwise stable equilibrium (PSNE) of the game then* G *is a Spiderweb graph with* $\max_{u \in K_r} t(u) \leq N(\varphi_p - \varphi_r(|V(K_r)| - 1) + 1)$ *and* $\forall r \in E_{peer} \setminus E_{K_r}$ *is a clear-cut peer edge (CPE), where CPE is a peer edge* $\overline{uv} \in G(s)$ *for which:*

- $\nexists w \in \mathcal{P} : v \in t(w) \wedge \overline{uw} \in G(s)$
- $\varphi_r < \min\{\frac{|t(u)|}{N}, \frac{|t(v)|}{N}\}$.

Finally the theorems lead to the following three corollaries.

Corollary 1. *In a PSNE a peer edge appears only if it is in* K_r *or both its endpoint ASs have sizable customer cones.*

Corollary 2. *For PSNEs there exists an upper bound for the size of the customer cones of the ASs in* K_r*, or more formally PSNE* $\implies \max_{u \in V(K_r)} t(u) \leq N(\varphi_p - \varphi_r(|V(K_r)| - 1) + 1)$.

Corollary 3. *In case of a PSNE there exists an upper bound for the size of* K_r *independent from* N*, i.e. PSNE* $\implies |V(K_r)| \leq \frac{\varphi_p + \varphi_r + 1 + \sqrt{(\varphi_p + \varphi_r + 1)^2 - 4\varphi_r}}{2\varphi_r}$

The above theorems deliver the following high-level sketch of the AS topology as a main intuitive message: (*i*) it is a Spiderweb-like graph with a clique (of tier-1 ASs) in the center and trees routed at the nodes of the clique, (*ii*) the peer edges appear more likely between ASs that have sizable customer cones, (*iii*) the size of the clique is constrained by the maintenance cost of peer and customer-provider edges and (*iv*) the largest customer cone size in the nodes of the clique is also driven by these maintenance costs.

4.2 A Game Theory-Based AS Level Internet Model (YEAS)

Using the analytical results of Section 4.1, in what follows I define a generative[5] AS topology model called YEAS that is able to create random topologies with similar statistical features. Such a model provides the possibility to furher analyze those statistical features that would be too complex to handle in the game theoretical framework. Besides recovering the usual features of network models (e.g. power-law degree distribution, large clustering, small diameter etc.) I implicitly encode the outcome of the analysis into the node and edge dynamics. Thus finally I require YEAS to produce Spiderweb-like graphs that have correct edge labeling, realistic tier-1 clique size and realistic placement of the peer edges. The framework of YEAS is based on the recently advocated hyperbolic space models presented in [20]. This basically dresses up a very simple hyperbolic model with the findings of Section 4.1.

[5] Generative here means that the created network is the result of a deterministic link creation process, in which the connectivity behavior of nodes is described by an algorithm (the Barabási-Albert model is a good example for generative models).

Thesis 2. *I have created a topology generator, called YEAS, that produces topologies based on the analyzed Nash Equilibria (e.g. Spiderweb graph). I have given proof that YEAS generated topologies have some implicit features like realistic power-law degree distribution, clustering coefficient, customer cone size distribution and peering likelihood. I have also compared the topologies with the AS level Internet measured by CAIDA and shown that YEAS outperforms a potpourri of existing models along several metrics.*

4.2.1 Topology Generation Process

Node layout The nodes are distributed (still representing the ASs) quasi-uniformly on the surface of a hyperbolic disk with radius R. This is done by assigning polar coordinates to each node as follows: $r = (1/\alpha)\,\mathrm{acosh}\,(1 + [\cosh(\alpha R) - 1]\, U_1)$ and $\phi = 2\pi U_2$, where U_1 and U_2 are independent random variables distributed uniformly over the interval $(0, 1)$ and α is a parameter controlling the heterogeneity of the layout.

Edge creation

To initialize take node u with the lowest radius and initialize a set $\mathcal{K} = \{u\}$[6]. In the first phase take nodes w one by one in an increasing order of their radii r_w and connect them to the others according to the following simple rule: if $Q \sum_{v \in \mathcal{K}} l(r_w, \phi_w, r_v, \phi_v) < \min_{v|r_v \leq r_w} l(r_w, \phi_w, r_v, \phi_v)$, then connect w to all nodes in $\mathcal{K}$ with peer edges and add w to $\mathcal{K}$, otherwise connect w to node $\mathrm{argmin}_{v|r_v \leq r_w} l(r_w, \phi_w, r_v, \phi_v)$ with a customer-provider edge. The constant Q is a tunable model parameter controlling the size of $\mathcal{K}$ and $l(r_u, \phi_u, r_v, \phi_v) = \mathrm{acosh}(\cosh r_u \cosh r_v - \sinh r_u \sinh r_v \cos(\phi_u - \phi_v))$. In the second phase every node $u \notin \mathcal{K}$ connects to a node v with a peer edge if $\nexists\, \overrightarrow{uv} \wedge l(r_u, \phi_u, r_v, \phi_v) < \varrho$, where ϱ is a parameter in the interval (0,R) for tuning peering willingness.

Thesis 2.1. *I have proved that topologies generated by YEAS have realistic power-law degree distribution and clustering coefficient.*

For proving the ability of YEAS to generate realistic power laws I show that the model generates all edges (u, v) for which $l(r_u, \phi_u, r_v, \phi_v) < \varrho$ but contains edges (u, v) for which $l(r_u, \phi_u, r_v, \phi_v) > \varrho$ with negligible probability.

In the case of the high clustering coefficient if we simply but arguably ignore (at least from the perspective of degree distribution and clustering) the negligible number of edges with length larger than ϱ we end up with a model readily analyzed in [20].

I have also compared YEAS generated topologies with the real AS level topology (derived from CAIDA's data). Fig. 7 shows the cumulative degree distribution of the real AS graph compared to the degree distribution of YEAS with setting $N = 40000$, $Q = 5$, $\alpha = 0.55$, $\varrho = 12.95$ and $R = 18.5$. The measured AS graph contains 41203 nodes, so I generated a similar sized topology. The clustering coefficient for the AS graph and for YEAS are both high 0.38 and 0.69, respectively.

Thesis 2.2. *I have analyzed YEAS generated topologies along the the expected customer cone size of the nodes and cone size distribution of the whole network. I have also compared the YEAS generated networks with CAIDA measurements along these metrics and found good matching.*

To analyze the average customer cone sizes I temporally omitted the peer edges generated by the model as these do not affect the customer cone sizes. Then the topology generation rules

[6] In YEAS this set represents the clique of tier-1 ASs.

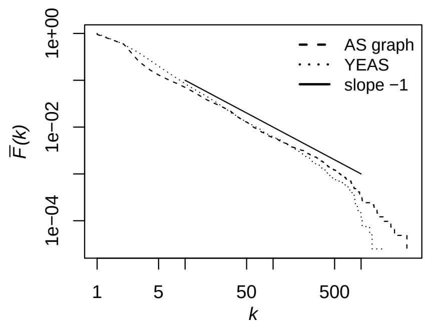


Figure 7. CCDF of degrees in the AS level and in the YEAS topology.

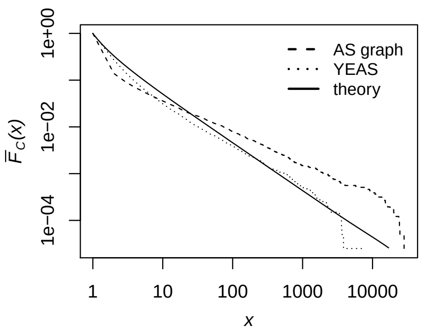


Figure 8. CCDF of customer cone sizes in the real AS graph, theory and YEAS topology.

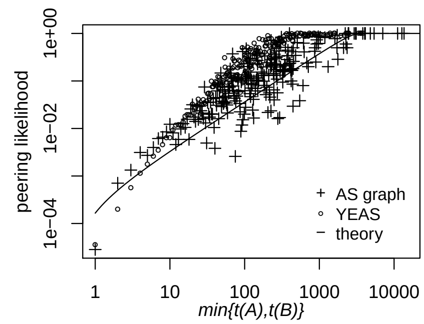


Figure 9. Peering likelihood between ASs as the function of their customer cone size.

Table 1. Comparison of a YEAS generated topology and CAIDA topology for basic metrics.

Network	Nodes	Edges	C. coef.	Avg. dist.	Avg. degree	Diameter	Max. cluster	# Tier-1
CAIDA top.	41203	116930	0.38	3.81	5.67	14	39327	16
YEAS	40000	115309	0.69	4.07	5.76	12	40000	16

can be translated into calculations in the hyperbolic geometry and with this I have been able to characterize the expected customer cone size and from that the cone size distribution. This result as the theoretical result goes hand in hand with the outcome of the simulations (Figure 8).

Thesis 2.3. *I have analyzed YEAS generated topologies along another quantity which is the peering likelihood and I have shown that it is in high correlation with the minimum of the customer cone sizes of the ASs. I have also compared the YEAS generated networks with CAIDA measurements along this metric.*

This means that the likelihood of peer edges of an AS that have a customer cone size to other ASs which have larger customer cone sizes is proportional to their cone size, and this likelihood tends to be 1, if the cone size is above a certain limit. This characteristic property is also confirmed by the simulations shown in Fig. 9 and coincides with results measured on the real AS topology.

Thesis 2.4. *I have compared YEAS generated topologies with CAIDA measurments and shown that it reflects the features of the AS level topology along several usual metrics. I have also compared it with a potpourri of existing models and shown that it outperforms most of them.*

Table 2. Comparison of network models.

Feature	Notation	Feature	Notation
Degree distr.	$\mathbb{P}$	Labeled	$\mathbb{L}$
Clustering	$\mathbb{C}$	Spider-like	$\mathbb{SL}$
Avg. distance	$\mathbb{D}$	Peering likelihood	$\mathbb{PL}$
Large size	$\mathbb{S}$	Few input params	$\mathbb{FP}$

		$\mathbb{P}$	$\mathbb{C}$	$\mathbb{D}$	$\mathbb{S}$	$\mathbb{L}$	$\mathbb{SL}$	$\mathbb{PL}$	$\mathbb{FP}$
C. oblivious	PLRG	✓	-	✓	✓	-	-	-	✓
	Inet	✓	-	✓	✓	-	-	-	✓
	dK-series	✓	✓	✓	✓	-	-	-	✓
C. aware	BA	✓	-	✓	✓	-	-	-	✓
	BRITE	✓	✓	✓	✓	-	-	-	✓
	SIMROT	✓	✓	✓	✓	✓	✓	✓	-
	H. et al.	✓	✓	✓	✓	-	-	-	-
	GENESIS	✓	✓	✓	-	✓	✓	✓	-
	YEAS	✓	✓	✓	✓	✓	✓	✓	✓

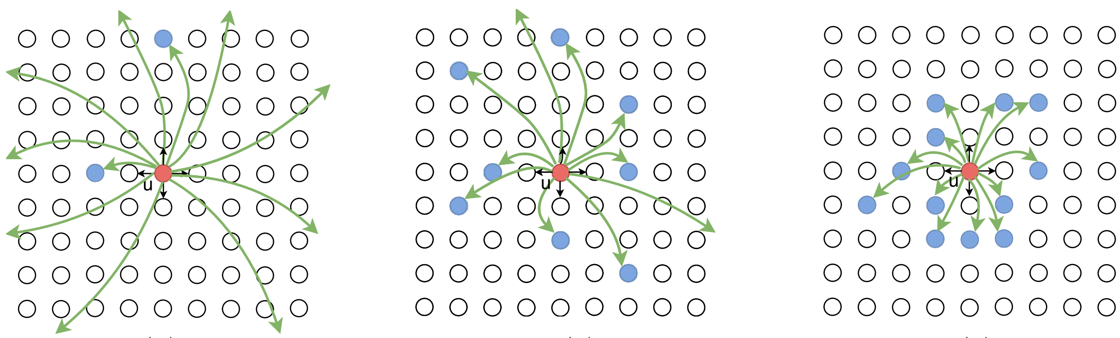
For the comparison I used YEAS with the same setting as previously ($N = 40000$, $Q = 5$, $\alpha = 0.55$, $\varrho = 12.95$ and $R = 18.5$). Table 1 provides an overview, it can be seen that the values of the various features are close to each other. Finally, I compared YEAS by outlining its features against a potpourri of existing models (Table 2). YEAS covers a wide range of feature set.

The above theoretical results show that YEAS generates realistic complex networks with proper degree distribution, clustering and diameter, yet incorporating the findings of Section 4.1 as the synthesized topologies are Spiderweb-like (trivially follows from the generation process), with tunable tier-1 clique (through the Q parameter) and realistic peering likelihood.

4.3 Topological Consequences of Greedy Navigation

Since Milgram's famous experiment [23] greedy navigability is a central issue in the theory of complex networks, as its great communication efficiency is confirmed in small words. A plausible explanation for the favorable navigational properties in such context is the assumed existence of a hidden metric space underneath these networks.

Kleinberg proposed an analytic model and a working algorithm that justifies the existence of such a *small-world* message forwarding [18] experienced by Milgram. In this work the world is modeled as a two-dimensional grid (Fig. 10), where each vertex is a person and there are local and long-distance edges between vertices. The probability of the existence of a long distance edge is $P_r(u,v) = d^{-r}(u,v)$, where d is the lattice distance and parameter $r \in [0,\infty)$. It is shown that a simple greedy routing algorithm needs $O(\log^2 n)$ time to travel between any pair of nodes if and only if $r = 2$ or more generally $r = D$, where D denotes the dimension of the lattice.



(a) $r << 2$ *(b) $r \sim 2$* *(c) $r >> 2$*

Figure 10. The effect of parameter r to the topology. Sub-figure a), b) and c) shows how likely longer connections emerge for different values of r. In the case of $r \sim 2$ most of the connections will lead to nodes relatively close to u, but there are some long connections as well, making possible of bypassing large distances, thereby enabling shortest paths, i.e. small-world property in the network.

Ever since the introduction of Kleinberg's lattice model [18] game theoretical investigation has been focused on explaining how such a network emerges due to the interaction of rational, selfish players. However, existing work assumes shortest path routing when measuring distance between nodes. There are several reasons why this view is limited, but the most important one is that since greedy routing is frequently used in both social and computer networks [4] to great success then it is worth to consider *"Why calculate the shortest path based equilibrium if players know they will route in a greedy manner?"*.

In the following I propose the Greedy Network Formation Game (GNFG) in which I assume a hidden metric space underneath the network and use the length of greedy paths as the measure of distance between players. Since shortest and greedy paths deviate in essence (see Figure 11), this shift will substantially change the corresponding equilibria.

Thesis 3. *I have defined a game theoretical model, which describes the communication in the complex networks more realistically, thus enable to better understand the formation process of*

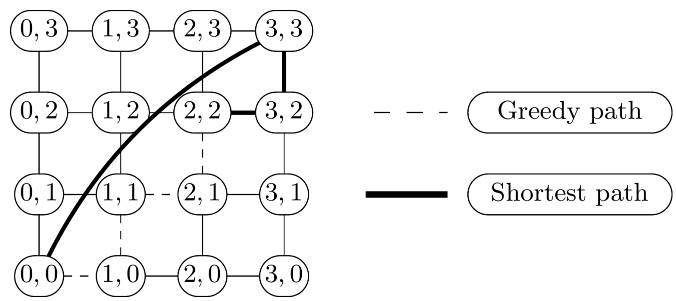


Figure 11. Deviation of shortest and greedy paths in the 2D Euclidean grid between nodes $(2,2)$ *and* $(0,0)$.

such networks. I have given a proof that the existence of greedy-routable small worlds cannot be economically justified under Kleinberg-like, constant dimensional, grid-based model. I have also shown that replacing Euclidean to hyperbolic space can lead to such Nash Equilibria for which the cost is better than the Social Optimum for Euclidean case.

Before introducing the game I recall the pioneering result of Kleinberg [18, 19] on greedy routing on Euclidean lattices, as it is used extensively in the arguments that are based on the analytical results.

Theorem 4. *(Kleinberg) Suppose that network nodes are placed in a 2-dimensional Euclidean lattice. From each node* u *one shortcut is added to the topology according to the distribution* $P(u,v) \sim l(u,v)^{-r}$*, where* $l(,u,v)$ *is the lattice distance between* u *and* v*. On this topology the expected delivery time of greedy routing is:*

$$E(t) = \begin{cases} C_1 \log^2(n) & \text{if} \quad r = 2, \\ C_2 n^{(2-r)/3} & \text{if} \quad 0 \le r < 2 \\ C_3 n^{(r-2)/(r-1)} & \text{if} \quad r > 2. \end{cases} ,$$

As Kleinberg states this result readily generalizes to lattices with higher dimensions.

4.3.1 The Greedy Network Formation Game

I define the Greedy Network Formation Game (GNFG) using Euclidean lattices, since the question is whether the Kleinberg-like grid network can emerge from the game.

Thesis 3.1. *I have defined the Greedy Network Formation Game and I have given the basic equilibrium states, which are the D-dimensional lattice and the full graph, depending on the link cost.*

Players, lattice and greedy routing – Let $\mathcal{P}$ be the set of players (identified with network nodes) with cardinality N. Players are placed into the vertices of a D-dimensional $\underbrace{n \times n \times \cdots \times n}_{D \text{ times}}$ lattice (i.e. n is the length of the lattice in each dimensions, so $n^D = N$), which is folded into a torus. The coordinate vector $\mathbf{u} = (u_1, u_2, \ldots, u_D)$ of player u indicates the position of u in the lattice. Distance between two players u and v used in the greedy routing decision is calculated as their lattice distance: $l(u,v) \overset{\text{def}}{=} l(\mathbf{u},\mathbf{v}) = \sum_{i=1}^{D} \min\{|u_i - v_i|, n - |u_i - v_i|\}$. A greedy routing step of player u operates over this metric space by choosing the neighbor whose lattice distance is the smallest from target t. If u has no neighbor v such that $l(u,t) > l(v,t)$ then greedy routing is in a local minimum and fails.

Strategies – A strategy for a node $u \in \mathcal{P}$ is to create a set of directed edges (arcs) to other nodes in the network; the strategy space is $S_u = 2^{\mathcal{P}\setminus\{u\}}$. Let s be a strategy vector: $s = (s_0, s_1 \dots s_{N-1}) \in (S_0, S_1 \dots S_{N-1})$ and $G(s)$ be the graph defined by the strategy vector s as $G(s) = \bigcup_{i=0}^{N-1}(i \times s_i)$. A mixed strategy is a probability distribution over the above (pure) strategies.

Payoff – The goal of the players is to minimize their cost function which is calculated as follows:

$$C_u(s) = \underbrace{\sum_{u \neq v} d_{G(s)}(u,v)}_{\text{communication cost}} + \underbrace{\varphi |s_u|}_{\text{link cost}}, \quad u, v \in \mathcal{P}, \tag{4}$$

where $d_{G(s)}(u,v)$ is the number of nodes involved in the greedy routing process between u and v (including v itself) over $G(s)$ and φ is the constant cost of creating one arc. By definition if greedy routing fails between u and v then $d_{G(s)}(u,v) = \infty$. This setting ensures that we get connected topologies in which there always exists a greedy path between any arbitrary pair of nodes.

Special cases for φ

I have characterized the equilibria of the game for special regions of φ: (i) if $1 < \varphi = O(N)$, any graph emerging from any NE or social optimum in the GNFG possesses the D-dimensional lattice as a subgraph, (ii) if $\varphi = \Omega\left(N^{1+1/D}\right)$ then the D dimensional lattice is a unique NE in GNFG and (iii) if $\varphi < 1$ then the full graph is a unique NE in the GNFG.

4.3.2 Simplified Greedy Network Formation Game

Deriving results for the GNFG in the region $1 < \varphi = O(N^{1+1/D})$ turns out to be a highly non-trivial problem. For the sake of tractability, in the following I restrict the argument to the one dimensional case and introduce the Simplified Greedy Network Formation Game (SGNFG). I will generalize the results later on. Any equilibrium or optimum solution of a Greedy Network Formation Game in one dimension always possesses the ring as a subgraph. Therefore I will play the SGNFG on a bi-directional ring, which implies that greedy routing will never fail. On this ring I define the SGNFG as follows: each player can create one directed edge only, which means that the strategy space reduces to a scalar e_u, which indicates the endpoint of the extra edge for player $u \in \mathcal{P}$. This also means that any player u will have a cost of $3\varphi < c_u < \infty$.

Social Optimum, Price of Anarchy and Price of Stability

Before the analysis it is important to mention that a NE is not necessarily optimal for players. In games with multiple equilibria, different equilibria can have (widely) different payoffs for the players. In order to be able to evaluate the different equilibria and to get a more precise picture about a game two distinguished metrics are defined, which are the *Price of Anarchy* (PoA) and the *Price of Stability* (PoS). In order to be able to define these precisely, first we need to define what the optimal outcome of a game is, that is called *Social Optimum* (SO).

Definition 5. *SO refers to an equilibrium state that maximizes the social welfare (i.e. minimizes the sum of all cost) even if its emergence requires a central coordination force (i.e. the independent* *decision-making is taken away from players). Formally, a strategy vector $s \in S$ constitutes a SO if:* $\sum_u I_u(s_u, s_{-u}) \geq \sum_u I_u(s'_u, s_{-u})$.

Based on the Social Optimum we can define PoA and Pos as follows:

Definition 6. *The PoA quantifies the loss to selfishness by comparing the performance at Nash equilibrium to the optimal state of a game. It is calculated as the ratio between the worst Nash equilibrium and the optimal outcome (SO). Formally:* $\frac{\min_{s\in\varepsilon}\sum_u I_u(s_u,s_{-u})}{\max_{s\in S}\sum_u I_u(s'_u,s_{-u})}$*, where ε is the set of equlibria.*

Definition 7. *The PoS is an optimistic form of the PoA as it shows how far the best-case scenario of the game, that is created by selfish players, lies from the optimum. Formally, the PoS of a game is the ratio between the best Nash equilibrium and the optimal outcome (SO):* $\frac{\max_{s\in\varepsilon}\sum_u I_u(s_u,s_{-u})}{\max_{s\in S}\sum_u I_u(s'_u,s_{-u})}$*, where ε is the set of equlibria.*

Based on these definitions I can characterize the different equilibria of the SGNFG.

Thesis 3.2. *I have defined the Simplified Greedy Game, that let us calculate equilibrium states in the region $1 < \varphi = O(N^{1+1/D})$. I have calculated the price of the Social Optimum $= O(N^2 \log^2(N)$, Price of Anarchy $= \Omega\left(\frac{N}{\log^2(N)}\right)$ and Price of Stability $= \Omega\left(\frac{N^{2/3}}{\log^2(N)}\right)$ and I have shown that there is not exist a Kleinberg-like solution in the game.*

When seeking for equilibrium solutions I will use mixed strategies, which means that the strategy of u is a random variable X indicating where to connect its extra edge. For the distribution $P(X = v) = p_v, p_u = p_{u-1} = p_{u+1} = 0$ and $\sum_{v\in\mathcal{P}} p_v = 1$ holds. Throughout analysis I - similarly as Kleinberg - assume that the distribution $P_X \in \mathbb{P}$ is decreasing and monotone, formally, $p_v \leq p_w$ if $l(u, v) > l(u, w)$ and $w \notin \{u - 1, u, u + 1\}$. This assumptions is fairly realistic, since otherwise the network does not bearing the properties of the underlying space and renders greedy routing meaningless. Let $A(u, v)$ denote the average number of greedy steps required to get from u to v.

Theorem 5. *The cost of the optimal solution of SGNFG is $O(N^2 \log^2(N))$.*

Price of anarchy in the SGNFG

Theorem 6. *The bi-directional Möbius ladder [14], in which the extra edges of the player are directed to exactly the opposite player on the one dimensional ring, is always a Nash equilibrium with total cost $\frac{N^3}{2}$. The Price of Anarchy in the SGNFG is therefore of $\Omega\left(\frac{N}{\log^2(N)}\right)$.*

Price of Stability in the SGNFG

To obtain equilibrium solutions the following two lemmas are needed. Let $A_e(u, v)$ denote the average number of greedy steps from u to v if player u has its extra edge connected to player e.

Lemma 1. *The larger the distance between two players, the more number of greedy steps is needed to travel between them on average. Formally: If $l(u, v) \leq l(u, w)$, then $A(u, v) \leq A(u, w)$ for $u, v, w \in \mathcal{P}$.*

Lemma 2. *If player u chooses a more distant player to connect its extra edge then the cost of u reduces. Formally: $\sum_{x\in\mathcal{P}} A_v(u, x) \geq \sum_{x\in\mathcal{P}} A_w(u, x)$, if $l(u, v) \leq l(u, w)$.*

From Lemma 1 and 2 one I can show, that the best strategy a player can have at any stage of the game is to uniformly choose among other players. The cost of player u is $c_u = \sum_{\forall v \neq u} A(u,v) = \sum_{\forall v \neq u} \sum_{j \in \mathcal{P} \setminus \{u-1,u,u+1\}} A_j(u,v) p_j$, which can be transformed to the form $c_u = \sum_{s \in S} p_s f(s)$.

Theorem 7. *If $f(s)$ is a monotonically decreasing function of $l(u, u{+}s)$ then in any given situation of SGNFG, player u's best response to the strategies of the other players is choosing the endpoint of its extra link uniformly. Formally:* $\mathrm{argmin}_{p \in \mathbb{P}} \sum_{s \in S} p_s f(s) = \mathrm{uniform}$.

Corollary 4. *The only Nash equilibrium of the SGNFG in mixed strategies is the case when all players connect their extra edge uniformly at random.*

Now that there is a clue for the structure of the network in equilibrium states, the cost of such equilibria can be calculated by borrowing again the results of Kleinberg.

Theorem 8. *The best Nash equilibrium of the SGNFG is of $\Omega(N^{8/3})$, therefore the Price of Stability is of $\Omega\left(\frac{N^{2/3}}{\log^2(N)}\right)$.*

According to [18] from the distributions of the form $p_v \sim l(u,v)^{-r}$, $r = D$ eventuates the only setting that produces a small-world topology, where the length of the greedy paths scales polylogarithmically with N. The conclusion from Corollary 4 is that $r = 0$ is the only possible setting to obtain a Nash equilibrium. This immediately leads to the following observation:

Proposition 1. *Kleinberg's optimal setting is not a Nash equilibrium, therefore small-world equilibrium solution does not exists for the SGNFG.*

4.3.3 Generalization of the Results

In the previous section I presented the in-depth analysis of the SGNFG and drew the negative conclusion that incorporating greedy routing within the network creation game takes the equilibrium topologies very far from the social optimum. What is more I have shown that a small-world network cannot be an equilibrium solution of the game. One might argue that the results may be valid only within the simple framework of SGNFG. Here I take a short look on the statements in more general settings of the game.

Thesis 3.3. *I have extended the Simplified Greedy Game to be able to introduce some more general settings, like multiple edges, distance-dependent link costs and multiple dimensions and I have shown that the Klienberg-like topology is still not an equilibrium state of the game.*

Multiple edges

In the simplified setting a player could have only one extra edge in addition to its lattice edges, however, in a general case a player can have multiple edges. Now I argue that if each player u can only afford a constant number of edges C_u then the equilibrium solution remains qualitatively the same. In the multiple edge case the cost of player u can be transformed to the form $c_u = \sum_{s \in S} p_s f(s)$ similarly to the single extra edge case (calculated from Lemma 1 and 2). Theorem 7 proves that the uniform distribution minimizes such cost functions. This also means that the best strategy that player u can have is to distribute its C_u edges uniformly in the lattice.

Distance-dependent link costs

In a general setting the cost of an edge may depend on the distance between its endpoints, which gives a more complex cost function

$$c_u = \sum_{v \in \mathcal{P}} p_v \left(\varphi(u,v) + \sum_{x \in \mathcal{P}} A_v(u,x) \right) = \sum_{s \in S} p_s f(s). \tag{5}$$

In this case however, $f(s)$ is not necessary monotonic, so I cannot prove that the uniform distribution is the only NE. What I can show is that a distribution which eventuates *strict* Nash equilibrium is uniform until a given lattice distance and zero otherwise.

Theorem 9. *If $p \in \mathbb{P}$ then $\exists f()$ for which p is weak Nash equilibrium. If $p \in \mathbb{P}$ is a strict Nash equilibrium, then $\exists r \in (0,1)$ that $p_s \in \{0, r\}$.*

Proposition 2. *A small-world topology can't be a strict Nash equilibrium.*

Multiple dimensions

For the sake of simplicity I carried out the proofs for the one dimensional case. In the following I illustrate that the argument can be extended to the finite D-dimensional case. First observe that the simple statement of Lemma 1 (the more distant a player is the more greedy steps are needed to travel between them) is the only result where the one dimensional assumption is exploited. Now I illustrate that Lemma 1 readily generalizes to higher dimensions.

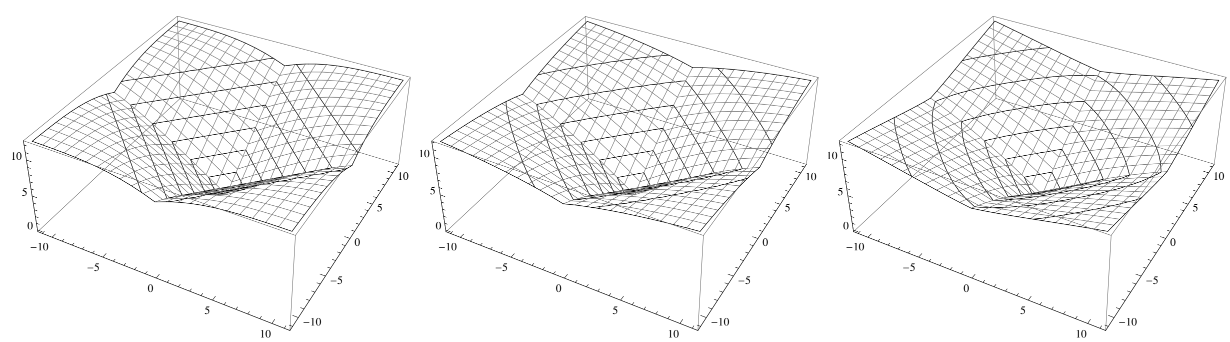


Figure 12. The average number of greedy steps ($A(u,v)$) between a reference player $u = (0,0)$ and the other players in the two dimensional lattice, if $p_v \sim l(u,v)^0$ (left), $p_v \sim l(u,v)^{-1}$ (center), $p_v \sim l(u,v)^{-2}$ (right).

Figure 12 shows the average number of greedy steps ($A(u,v)$) required to travel between a reference player (at the center of the figure) and the other players in the two dimensional lattice. If w denotes the neighbor of u who is closest to v, then $A(u,v)$ can be calculated by following recursion:

$$A(u,v) = \sum_{x \in \mathcal{P}} p_x A_x(u,v) = \sum_{x: l(x,v) < l(w,v)} p_x (1 + A(x,v)) + \left(1 - \sum_{x: l(x,v) < l(w,v)} p_x \right) (1 + A(w,v)).$$

Figure 12 supports the conjecture that $A(u,v)$ grows with the lattice distance if the game is played in multiple dimensions.

Conjecture 1. *Small-world topologies cannot emerge as equilibria from the SGNFG even if the dimension of the lattice is raised to an arbitrary constant value. This means that the existence of small-worlds cannot be economically justified under the Kleinberg-like constant dimensional grid-based models.*

4.3.4 Hyperbolic Space

The results support the claim that small-world networks cannot be equilibrium solutions of the Greedy Network Formation Game even if the game is played under fairly generalized conditions. So the question arises: *"How can small-world topologies emerge?"* In the hyperbolic space I prove that socially optimal solutions can readily emerge from the Greedy Network Formation Game.

Thesis 3.4. *I have proved that by replacing Euclidean to hyperbolic space then any emerging Nash Equilibrium contains a tessellation for that the cost is better than the Social Optimum for Euclidean case, which is $O(N^2 log(N))$ and $O(N^2 log^2(N))$, respectively.*

For the investigation I use the Poincaré disk model [2] of the two dimensional hyperbolic space as this model makes the calculations easier here. In this space player u has a coordinate vector $\mathbf{u} = u_1, u_2 \in [0, 1)$ and the distance between u and v is calculated according to the Poincaré distance function:

$$d_p(u, v) \stackrel{\text{def}}{=} d_p(\mathbf{u}, \mathbf{v}) = \operatorname{arccosh}\left(1 + 2\frac{||\mathbf{u} - \mathbf{v}||^2}{(1 - ||\mathbf{u}||^2)(1 - ||\mathbf{v}||^2)}\right),$$

where $||x||$ stands for the Euclidean norm of x.

The players are placed at equal distances from each other similarly to the case of the two dimensional Euclidean lattice, thus the players will be located in the vertices of a so-called hyperbolic tessellation (see Figure 13). A tessellation [17] can be characterized by a pair (ν, κ) where ν stands for the vertex number of its constituent polygons and κ denotes the number of meeting polygons at a given vertex. For (ν, κ), $\frac{1}{\nu} + \frac{1}{\kappa} < \frac{1}{2}$ must hold. A graph $T(V, E)$ can be constructed from the tessellation if its vertices are considered as the vertices of the graph and the sides of the polygons as edges.

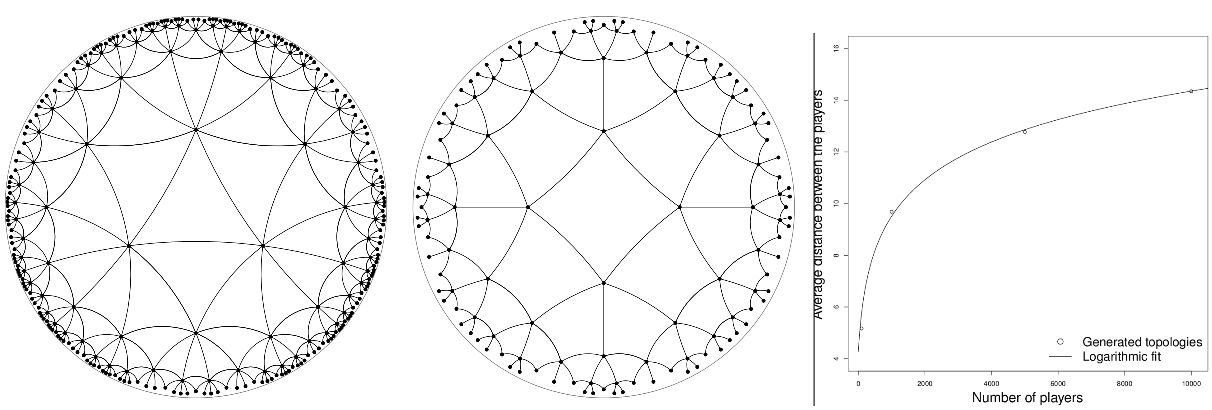


Figure 13. (3,8) (left) and (4,5) (middle) hyperbolic tessellations and the average distance between the vertices in the (4,5) hyperbolic tessellation as a function of the number of vertices (right).

In this setting of GNFG, similarly to the Euclidean game the following lemma holds.

Lemma 3. *Any graph emerging from any Nash equilibrium or Social Optimum in the two dimensional hyperbolic GNFG possesses the underlying tessellation graph T as a subgraph.*

Lemma 4. *The number of vertices in a (ν, κ) tessellation grows exponentially with the number of layers, thus the diameter of the tessellation graph T is of $O(\log N)$.*

From Lemma 4 we can see that using only the edges of the tessellation (without any extra shortcut edges) player v can be reached from u in $O(\log(N))$ number of greedy steps (see Figure 13). The total cost of the GNFG is therefore of $O(N^2 \log(N))$, which is better than the Social Sptimum for Euclidean case.

5 Applications

To conclude the dissertation here I discuss how the obtained results can be used to further extend our knowledge about the topology of complex networks. In this chapter I list some application scenarios for each chapter where results are introduced.

Topological Consequences of the BGP routing policy

Internet specific knowledge can greatly help to improve the performance of the network. The more insight we gather on how BGP drives the topology formation of the Internet's AS level network the easier it is (*i*) to design better routing policies, (*ii*) to understand why and how the traffic emerges and (*iii*) to optimize the current network structure. The most specific example is clearly the area of Content Delivery Networks (CDN) [24], where global topological peculiarities are highly exploited e.g. in surrogate and cache placement strategies or request routing mechanisms. Note that CDN is just a narrow segment of the whole spectrum. To give a few more examples, the placement of data centers [12], peer-to-peer networks [8, 21], traffic engineering [3], business based AS peering strategies [9] can also largely benefit from Internet topology related knowledge. The investigation of the AS topology is also a popular topic in the network science community that consolidates researchers from diverse or multidisciplinary research areas [5, 6, 1, 7, 26, 22, 25].

A Game Theory-Based AS Level Model (YEAS)

Topology generators are often used in diverse testing processes of different applications. Testing of novel routing policies, traffic handling or security algorithms requires realistic topologies that have some randomness but bear similar statistical features to real networks at the same time. Certainly the needs are not exactly the same in all cases, so the topology generators can be categorized along the needs they serve. YEAS can be useful in situations when quickly generated large topologies with the characteristics of the Internet's AS level network are needed, including labeled nodes and connections according to business considerations.

Topological Consequences of the Greedy Navigation

Greedy routing is the most accepted policy in describing the communication process of real-world complex networks. We have empirical evidences (e.g. the Milgram experiment) that in many cases this method enables efficient information distribution among network nodes. However, creating a game that explains the emerging process of such networks had been a non-trivial issue in game theory until recently. My results are cited in two papers, by Yang et. al [27] and Gulyás et. al [13], where the authors manage to explain the emergence of such networks.

References

[1] Réka Albert, Hawoong Jeong, and Albert-Lászlo Barabási. Error and attack tolerance of complex networks. *Nature*, 406(6794):378–382, 2000.

[2] J.W. Anderson. *Hyperbolic geometry.* 2005.

[3] Daniel Awduche et al. Overview and principles of Internet traffic engineering. Technical report, RFC 3272, May, 2002.

[4] Mm Boguñá and D. Krioukov. Navigating ultrasmall worlds in ultrashort time. *Physical Review Letters*, 102(5):058701, Feb 2009.

[5] M. Boguna, D. Krioukov, and K. C. Claffy. Navigability of complex networks. *Nature Physics*, 5(1):74–80, 2009.

[6] M. Boguna, F. Papadopoulos, and D. Krioukov. Sustaining the Internet with hyperbolic mapping. *Nature Communications*, 1(6):1–8, 2010.

[7] Claudio Castellano and Romualdo Pastor-Satorras. Competing activation mechanisms in epidemics on networks. *Scientific Reports*, 2, 2012.

[8] Miguel Castro, Peter Druschel, Y Charlie Hu, and Antony Rowstron. Topology-aware routing in structured peer-to-peer overlay networks. In *Future Directions in Distributed Computing*, pages 103–107. Springer, 2003.

[9] David Clark, William Lehr, and Steven Bauer. Interconnection in the Internet: the policy challenge. In *Research Conference on Communication, Information and Internet Policy*, 2011.

[10] L. Gao and J. Rexford. Stable internet routing without global coordination. *IEEE/ACM Transactions on Networking*, 9(6):681–692, 2001.

[11] Lixin Gao. On inferring autonomous system relationships in the internet. *IEEE/ACM Transactions on Networking (ToN)*, 9(6):733–745, 2001.

[12] Albert Greenberg et al. The cost of a cloud: research problems in data center networks. *ACM SIGCOMM CCR*, 39(1):68–73, 2008.

[13] András Gulyás, József Bíró, Attila Kőrösi, Gábor Rétvári, and Dmitri Krioukov. Navigable networks as Nash equilibria of navigation games. *Nature Communications*, Vol. 6, 2015.

[14] R.K. Guy and F. Harary. On the Möbius ladders. *Canadian Mathematical Bulletin*, 10:493–496, 1967.

[15] G. Huston. Interconnection, peering and settlements: Part I. *Internet Protocol Journal*, 2(1), JUN 1999.

[16] Matthew O Jackson. A survey of network formation models: stability and efficiency. *Group Formation in Economics: Networks, Clubs, and Coalitions*, pages 11–49, 2005.

[17] D.E. Joyce. Hyperbolic tessellations. *Clark University. http://aleph0. clarku. edu/˜ djoyce/poincare/poincare. html. Updated December*, 1998.

[18] J. Kleinberg. The small-world phenomenon: an algorithm perspective. In *Proceedings of STOC'00*, pages 163–170, 2000.

[19] J. Kleinberg. Complex networks and decentralized search algorithms. In *Proceedings of ICM*, volume 3, pages 1019–1044. Citeseer, 2006.

[20] D. Krioukov et. al. Hyperbolic geometry of complex networks. *Physical Review E*, 82(3):036106, 2010.

[21] Eng Keong Lua, Jon Crowcroft, Marcelo Pias, Ravi Sharma, Steven Lim, et al. A survey and comparison of peer-to-peer overlay network schemes. *IEEE Communications Surveys and Tutorials*, 7(1-4):72–93, 2005.

[22] Sergei Maslov, Kim Sneppen, and Alexei Zaliznyak. Detection of topological patterns in complex networks: correlation profile of the internet. *Physica A: Statistical Mechanics and its Applications*, 333:529–540, 2004.

[23] Stanley Milgram. The small world problem. *Psychology Today*, 2(1):60–67, 1967.

[24] Mukaddim Pathan and Rajkumar Buyya. *Content Delivery Networks*, chapter A Taxonomy of CDNs, pages 33–77. Springer Berlin Heidelberg, Berlin, Heidelberg, 2008.

[25] V Rosato, L Issacharoff, S Meloni, D Caligiore, and F Tiriticco. Is the topology of the Internet network really fit to sustain its function? *Physica A: Statistical Mechanics and its Applications*, 387(7):1689–1704, 2008.

[26] Bo Xiao, Lian-dong Liu, Xiao-chen Guo, and Ke Xu. Modeling the ipv6 internet as-level topology. *Physica A: Statistical Mechanics and its Applications*, 388(4):529–540, 2009.

[27] Zhi Yang and Wei Chen. A game theoretic model for the formation of navigable small-world networks. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1329–1339. ACM, 2015.

References

Journal papers (11.33 p)

[J1] David Szabo, Andras Gulyas. Notes on the Topological Consequences of BGP Policy Routing on the Internet AS Topology. *Lecture Notes in Computer Science Volume 8115, 2013, pp 274-281.* **(6 p)**

[J2] **Dávid Szabó**, Attila Kőrösi, József Bíró, András Gulyás A Deductive Way of Reasoning about the Internet AS Level Topology. *Chinese Physics B.*
(6/3=2 p)

[J3] Gulyás András, Kőrösi Attila, **Szabó Dávid**, Biczók Gergely. On greedy network formation. *SIGMETRICS PERFORMANCE EVALUATION REVIEW*, 40:pp. 49-52, 2012. **(6/3=2 p)**

[J4] **Szabó Dávid**, Gulyás András, Csernai Márton, Heszberger Zalán. Struktúrafüggetlen címzésen alapuló önszerveződő útvonalválasztási architektúra. *HÍRADÁSTECHNIKA LXVI. évfolyam*, 2011/3: pp. 2-10, 2011. **(4/3=1.33 p)**

Conference papers (4.25 p)

[C1] **Szabó Dávid**, Gulyás András. Jelterjedési utak topológiára gyakorolt hatásának vizsgálata játékelméleti módszerekkel. Mesterpróba konferencia, 1. helyezés, 2012. **(1 p)**

[C2] **David Szabo**, Andras Gulyas, Frank H.P. Fitzek, Daniel E. Lucani. Towards the Tactile Internet: Decreasing Communication Latency with Network Coding and Software Defined Networking. European Wireless 2015. **(3/3=1 p)**

[C3] **David Szabo**, Andras Gulyas, Balazs Sonkoly, Felician Nemeth, Frank H.P. Fitzek. Towards the 5G Revolution: A Software Defined Network Architecture Exploiting Network Coding as a Service. SIGCOMM 2015.

[C4] András Gulyás, Attila Kőrösi, Gábor Rétvári, József Bíró, **Dávid Szabó**. Brief Announcement: Network Formation Games Can Give Rise to Realistic Networks. In: 31st Annual ACM SIGACT-SIGOPS Symposium on Principles of Distributed Computing. Madeira, Portugália, 2012.

[C5] **Dávid Szabó**, Csernai Márton. Self organizing structurless routing architecture. In: POSTER 2011: 15th International Student Conference on Electrical Engineering. Prága, Csehország, 2011. **(3/2=1.5 p)**

[C6] Gábor Rétvári, **Dávid Szabó**, András Gulyás, Attila Kőrösi, János Tapolcai. An information-theoretic approach to routing scalability. ACM Hotnets 2014. **(3/4=0.75 p)**

Independent Citations

[J1-1] Lenzner, Pascal. Greedy selfish network creation. Internet and Network Economics. Springer Berlin Heidelberg, 2012. 142-155.

[J1-2] Yang, Zhi, and Wei Chen. A Game Theoretic Model for the Formation of Navigable Small-World Networks. Proceedings of the 24th International Conference on World Wide Web. International World Wide Web Conferences Steering Committee, 2015.

[J1-3] Zolanvari, Maede. SDN for 5G. CSE570S: Recent Advances in Networking, available: http://www.cse.wustl.edu/ jain/cse570-15/ftp/sdnfor5g.pdf.

[J1-4] Do-Duy, Tan, and M. A. Vazquez-Castro. Network Coding function virtualization. Signal Processing Advances in Wireless Communications (SPAWC), 2016 IEEE 17th International Workshop on. IEEE, 2016.

[J1-5] Bouzghiba, Soukaina, et al. Towards an Autonomic Approach for Software Defined Networks: An Overview.

[J1-6] Maier, Martin, et al. The tactile internet: vision, recent progress, and open challenges. IEEE Communications Magazine 54.5 (2016): 138-145.

[J1-7] Heinonen, Johanna, et al. Mobility management enhancements for 5G low latency services. Communications Workshops (ICC), 2016 IEEE International Conference on. IEEE, 2016.

[J1-8] Vukobratovic, Dejan, et al. CONDENSE: A Reconfigurable Knowledge Acquisition Architecture for Future 5G IoT. IEEE Access 4 (2016): 3360-3378.

[J1-9] Confidential, C. O. OpenLab: Extending FIRE testbeds and tools, FP7 EU project Deliverable D2. 1 Experimental plane–Experiment Controllers.