GENERATION OF ROBUST NETWORKS WITH OPTIMIZATION UNDER BUDGET CONSTRAINTS

László Gulyás

Computation and Automation Research Institute
Hungarian Academy of Sciences
Budapest, H-1111 Kende u. 13-17
gulya@omega.ailab.sztaki.hu

ABSTRACT

This paper addresses the problem of generating networks that are robust against random failures (i.e., against the random removal of nodes). We construct an agent-based model in which agents represent the nodes of the network that connect to one another aiming to maximize their connectivity. Each agent can build a fixed number of links. However, information about the existing network is costly, so the agents must optimize under budget constraints, i.e., only having information about a limited number of existing nodes. Numerical simulation shows that this scheme generates robust networks under a wide range of conditions. A key observation is that the pattern of information access, determined by the scheme used for pricing information about the existing network, is pivotal for the desired system-level property.

Keywords: networks, robustness, budget constraints, optimization, agent-based simulation

1. INTRODUCTION

The various networks forming what we generally call 'the Internet' tend to share an interesting statistical property. The networks' degree distribution follows a power law. Such networks are often referred to as 'scale-free networks', as they don't have a characteristic scale (they are self-similar, independent of the scale). In the case of Internet, this is true at various levels. It doesn't matter if one looks at routers and their connections, at inter-domain links, or the documents and links of the World Wide Web; the network is scale-free. (Albert and Barabási, 2002) (Faloutsos *et al*., 1999)

It is likely that this scale-free property of Internet's networks made an important practical contribution to the Internet's success. (Albert and Barabási, 2002) Namely, power law networks are robust against the random removal of the nodes (i.e., the failure of the given node). From this point of view, the goodness of the network depends on its connectivity, measured by the number of nodes in its largest connected component. In power law networks, the expected connectivity remains high even after a series of random failures. Clearly, the power law degree distribution implies that a uniformly selected random node will be likely to have only a few links. Thus, its removal cuts off only a few nodes from the largest component. (See Figure 1.)

The downside of the power law degree distribution is that it makes the network vulnerable to deliberate, planned attacks. Obviously, if an attacker gradually removes the few extremely connected nodes, the network soon disintegrates. Nonetheless, robustness makes power law distribution an often-sought property of engineered networks.

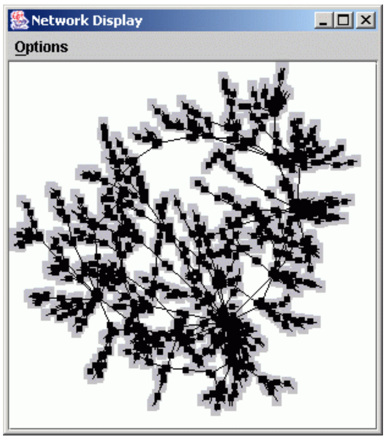


Figure 1: A network[i] with power law degree distribution. A few highly connected 'hubs', and many 'connectors' that are only linked to one of the hubs characterize such 'scale-free networks'.

This paper addresses the problem of generating networks that are robust against random failures (i.e., against the random removal of nodes), subject to certain practical constraints:

- *Network generation is incremental.*
  Nodes become available at different moments in time. Therefore, the generation algorithm must be able to link new nodes to the existing network, preserving its robustness. This ensures the extensibility of the network. For the sake of simplicity it is assumed that at most one node is created at any given moment.
- *Information about the existing network is costly.*
  When determining a new node's links, the generation algorithm must assess the existing network. In large real-world networks, it is unrealistic to assume that exact, up-to-date information is readily available about the degree of a given node. Rather, this information must be collected via computationally intensive processes.

We investigate this problem via numerical simulation of an agent-based model. Arriving nodes are represented as agents that must select a fixed number of nodes (the same for all agents) to link to. It is assumed that the agents want to maximize the number of nodes reachable via their links. They attempt to achieve this by linking to the nodes with the highest degree. Once an agent created its links, it becomes passive. It can only receive further links if it is selected by subsequent agents. Initially the agents have no knowledge about the existing network. They must buy information about the degree of older nodes from a central authority. However, agents have budget constraints, so they can only inspect a limited number of existing nodes. Therefore, they will link to the best-connected nodes among the inspected ones. The price of the information about the number of links of a given node is independent of the node in question, but may depend on the size of the network according to a pricing scheme.

As the agents have no previous knowledge about the network, their requests to the central authority cannot specify the older node they are interested in. At most, they can list the nodes they already have knowledge about. In response to such a list, the central authority returns a random node not contained by the list, together with its degree.

The paper is structured as follows. The next section presents the detailed model and the investigated pricing schemes. Section 3 summarizes the results gained by numerical simulation. This is followed by the discussion of related and future works. Finally, Section 5 concludes the paper.

2. THE MODEL

2. 1 The Base Model

Let's identify the agents (nodes) by natural numbers ($i \in$ N), the number denoting the time step (iteration) at which the given node joins the network. Let A^t=[1, t] denote the set of agents forming the network at time step t. Similarly, $link^t \subseteq A^t \times A^t$ denotes the directed links of the network generated so far. (In the following, index t will be omitted, whenever the meaning can be unambiguously determined from the context.) The outward edges of agent i are given by out_i=Image($link(i)$). Similarly, in_i stands for the number of links pointing to the i^{th} node.

It is assumed that out_i=E for all agents, where E is a positive integer and a parameter of the model. The network generation process starts with E fully connected nodes. (For the purposes of indexing it is assumed that these agents arrived at time steps 1, 2, …, and E.)

When subsequent agent j arrives, it buys information about the maximum number of existing nodes allowed by its budget constraint $b_j \in$ R$^+$ (another set of parameters to the model). The price of one such piece of information is determined by the used pricing scheme, PS: N $\rightarrow$ R$^+$. The value $PS(i)$ determines the price that an agent has to pay to receive information about a node's connectivity in a network composed of i agents. Therefore, the number of nodes agent j buys information about is

$$d_j = \min\left(j-1, \max\left(E, \lfloor b_j / PS(j-1) \rfloor\right)\right) \qquad (1)$$

Thus, agent j receives information about the connectivity of d_j nodes randomly selected by the central authority (without replacement). Let's denote this set by $C_j \subseteq A^t$. Agent j then selects the E most connected nodes, S_E, from C_j and creates an outgoing link to each of them.

$$S_E = \left\{ a \in C_j, \quad \left| \left\{ i \in C_j, \quad in_i + out_i > in_a + out_a \right\} \right| < E \right\} \qquad (2)$$

If the selection of S_E is not unambiguous (i.e., there are several inspected nodes with the same number of links), uniform random selection is applied.

2.2 Budget Constraints and Pricing Schemes

We consider two classes of budget constraint definitions. In the homogenous case, all agents have the same budget, i.e., b_i=B for all i. On the other hand, in the heterogeneous case b_i's are distributed uniformly in the interval [1, B]. (In both cases, $B \in$ N is a parameter of the model.)

We also consider three pricing schemes: one in which price is independent of, one in which cost grows and another in which it decreases with the size of the network. The decreasing pricing scheme can be motivated by the possibility of a positive 'economies of scale'. That is, in a growing network, it may become cost-effective for the central authority to maintain a database of answers to previous questions. The exact choices for the PS function are the following:

PS1:	$PS(i) = C$,	(3)
PS2:	$PS(i) = C*B / i$, and	(4)

PS3: $PS(i) = i \,/\, C$, (5)

where $C \in N$ is a model parameter.

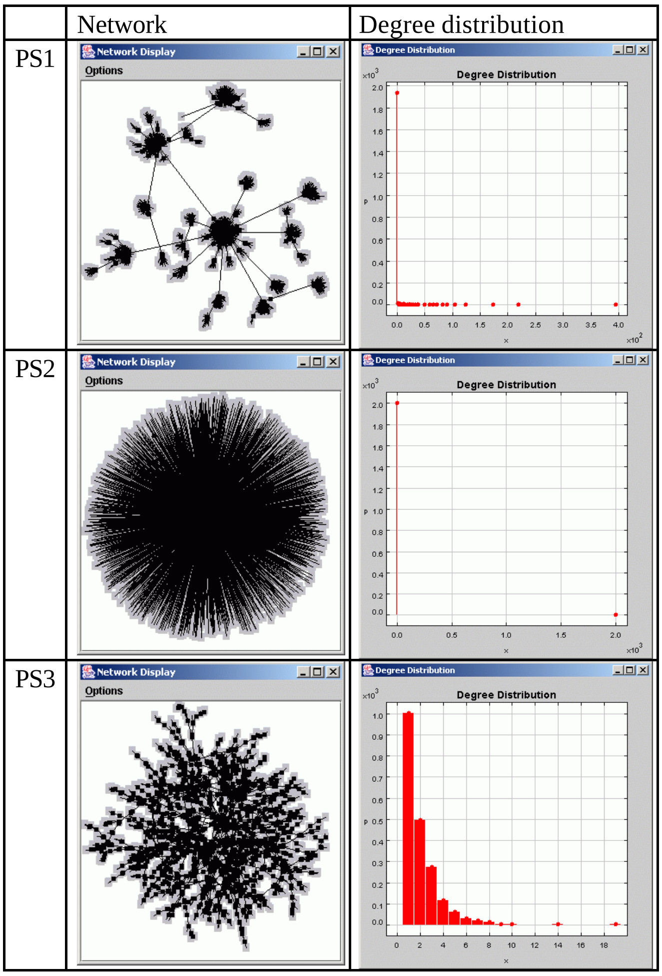


Figure 2: The network and its degree distribution after about 2000 iterations with homogenous budget constraints for the three studied pricing schemes. (E=1, B=100, C=1.)

3. RESULTS

3.1 Robust Networks

In this section we summarize the results of our experiments with the various budget constraints and pricing scheme combinations. These results were obtained via numerical simulations[ii] that were performed using the RePast agent-based simulation framework. (Cederman and Gulyás, 2001) (RePast, 2003) Our key finding is that the model generates robust networks under a wide range of conditions. Various combinations of the above pricing schemes and budget constraints yield robust networks.

Figure 2 and Figure 3 show examples from our explorations with different initial configurations. They depict the networks and their degree distribution after the arrival of approximately 2000 agents (nodes) for the three pricing schemes investigated. On the horizontal axis of the degree distribution figures the number of edges ($in_i + out_i$) is measured, while the vertical axis depicts the number of nodes (agents) having the given degree.

It is clear from the figures that all three pricing schemes lead to networks where nodes with a low degree are over-represented. However, it is also clear that this bias is stronger in case of PS1 and PS2. It is also worth noting that the networks generated under homogeneous budget constraints are qualitatively similar to those yielded by the heterogeneity of the agents' financial means. The only exception here is PS2, the decreasing pricing scheme, where a homogenous constraint leads to a 'star topology'. In such a network all nodes but the initial ones are connected to a single core. This extreme bias for low degree nodes makes the network very robust against random failures, but it is often undesirable for other practical reasons. The reason for the emergence of the star topology is simple. In this setup each agent can buy information about the entire network, so it is always the most connected node that is selected. On the other hand, it is easy to see that the node receiving the first inward link will have an advantage that others can never make up for.

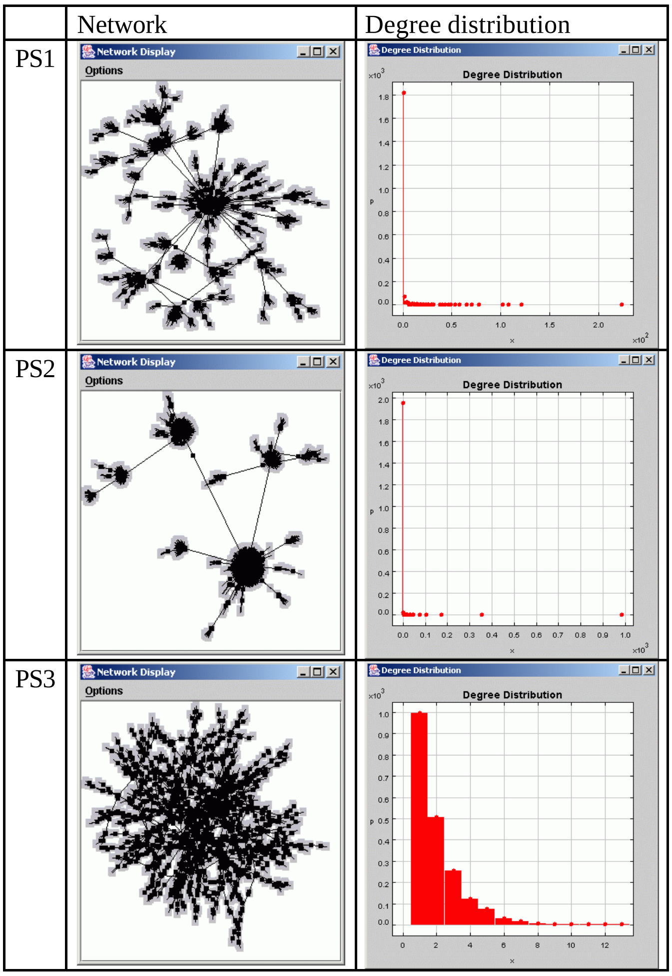


Figure 3: The network and its degree distribution after about 2000 iterations with heterogeneous budget constraints for the three studied pricing schemes. (E=1, B=100, C=1.)

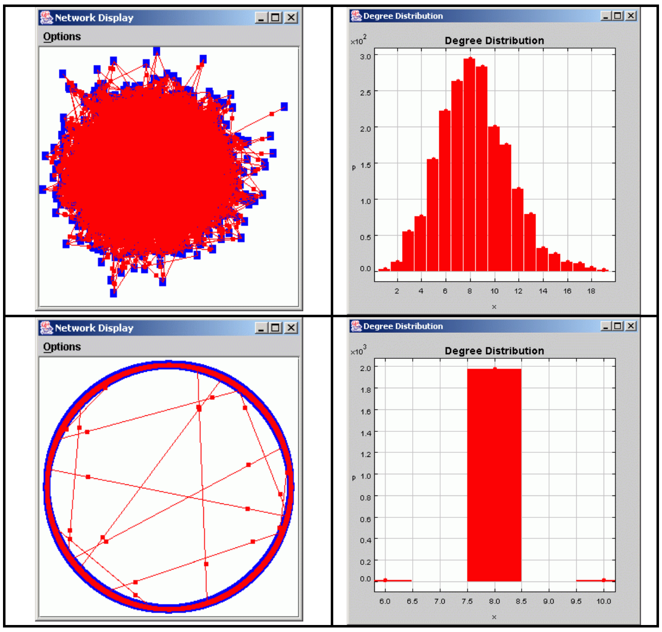


Figure 4: The network view and the degree distribution of an Erdos-Rényi network (top row), and a Watts-Strogatz network (bottom row). In both cases, the network consists of 2000 nodes and the parameters are chosen to ensure that the average degree per node is eight. (In case of the ER network we used *p*=0.0021, while in case of the WS network the parameters were *k*=2 and *p*=0.001.)

Figure 2 and Figure 3 made it clear that our model produces networks that are biased towards low degree nodes. It is not clear, however, how these networks differ from those generated by other, better known and more thoroughly studied network models. To investigate this issue, we have compared our networks to those created by the Erdos-Rényi (random density), and Watts-Strogatz ('small-world') network models. (Erdos and Rényi, 1959) (Watts and Strogatz, 1998) Figure 4 shows the views used earlier applied to examples from these experiments. In the case of both models, the parameters were chosen to ensure that the basic network statistics are comparable to those produced by our model. For the networks presented here, the number of nodes was fixed at 2000 in all cases, and the average number of links was set to 8.

Figure 5 shows the degree distributions of the networks produced by our model. Comparing it to Figure 4 makes the difference obvious, especially in case of pricing schemes PS1 and PS2. It is clear that the model introduced in this paper produces networks whose degree distribution is significantly more biased towards low degree nodes than those produced by the two most commonly used network generation models, namely those of Erdos and Rényi (1959) and Watts and Strogatz (1998). As the over-representation of low degree nodes implies that a random selection will be likely to return a node with a low number of links, such networks tend to be more robust against random failures of their nodes.

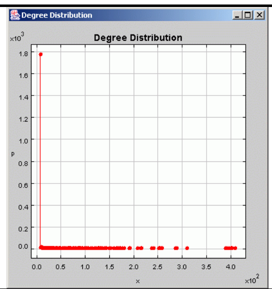
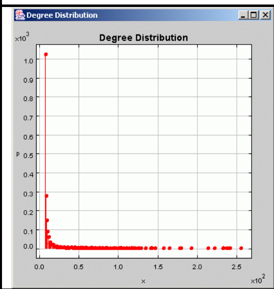
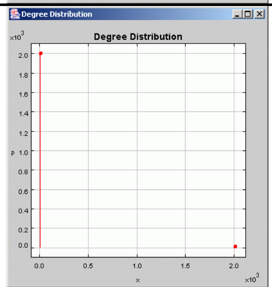
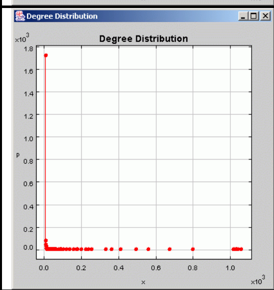
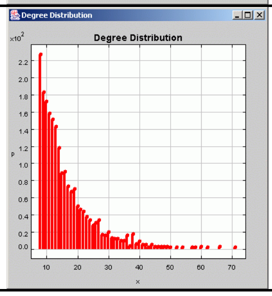
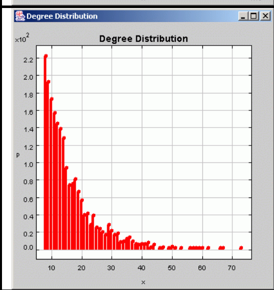
	Homogeneous Budget Constraints	Heterogeneous Budget Constraints
PS1		
PS2		
PS3		

Figure 5: The degree distribution after about 2000 iterations for the three pricing schemes with both homogenous and heterogeneous budget constraints. (E=8, B=100, C=1.)

3.2 Special Topologies

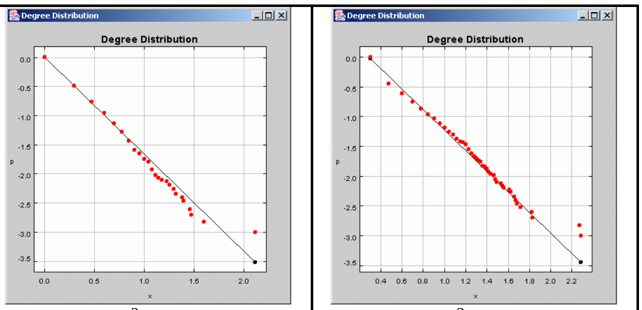
In addition to the generation of robust networks, the model also yields special topologies under certain parameter combinations. We discussed the 'star topology' in the previous section. However, the model is also capable of generating another special topology, namely 'scale-free networks'.

The often sought-for topology that is called a scale-free network is characterized by a degree distribution that follows a power law. (Newman, 2003) That is, if p_k stands for the frequency of nodes with degree k then

$$p_k = k^{-a} \tag{6}$$

where $a \in R$ is a constant. Typically, power law distributions are depicted on a log-log graph (a logarithmic scale is applied to both axes), where they appear as a straight (usually downward sloping) line. Taking the logarithm of both sides of Equation (6), one gets:

$$\log p_k = -a \cdot \log k + b \qquad (7)$$



a) (R^2=0.997) **b)** (R^2=0.988)

Figure 6: Power law degree distributions with PS2. The graphs apply a logarithmic scale on both axes. The continuous lines show the results of linear regression. The network on the left hand side (a) was generated using heterogeneous, while the one on the right (b) using homogeneous budget constraints. Both networks contain 2000 nodes. (The exact model settings were: a) E=1, B=100, C=100, and b) E=2, B=100, C=100.)

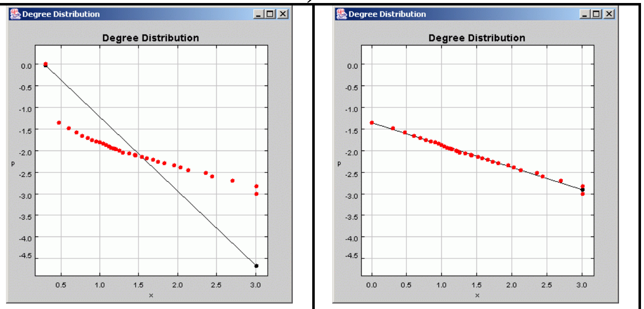
This relationship is often used to test whether empirical or computer generated distributions follow a power law. The method is to apply linear regression to the log-log transformation of the distribution, and check the goodness-of-fit (R^2). This value is, by definition, between 0 and 1, and the closer it gets to 1, the better the fit is. That is, the more closely the distribution follows a power law.

The model discussed in this paper generates networks with power law degree distributions under several circumstances. PS2, the particular formulation of the increasing pricing scheme, yields d_i's as a hyperbolic function of the number of nodes:

$$d_i = \begin{cases} \frac{BC}{i-1}, & i-1 < BC \\ E, & i-1 \geq BC \end{cases} \qquad (8)$$

This leads to scale-free networks under both homogenous and heterogeneous budget constraints, as shown on Figure 6, if the system size is not significantly greater than BC.

On the other hand, the decreasing pricing scheme (PS1) under heterogeneous budget constraints is also capable of generating power law distributions. When the agents' information access is distributed uniformly between knowing the full network and being able to inspect a single node only, the resulting network will have a distribution that is close to a power law (see Figure 7a). Unfortunately, these networks have a little too high bias for nodes with the lowest possible degree (i.e., E). Closer inspection reveals that this is due to nodes that never receive links after joining the network. Indeed, as shown on Figure 7b, the in-degree distribution, one that is concerned with inward edges only, shows a perfect power law for the largest connected component (containing all the nodes, but the ones with no inward links).



a) (R2=0.748) **b)** (R2=0.997)

Figure 7: Power law degree distribution with PS1. The graphs apply a logarithmic scale on both axes. The continuous lines show the results of linear regression. Both graphs depict the same 2000 node network that was generated under heterogeneous budget constraints, with E=2, B=200, and C=1. The one on the left hand side (a) shows the distribution of in_i+out_i, while the one on the right (b) the distribution of in-degrees, i.e., that of in_i.

4. RELATED AND FUTURE WORK

In recent years scale-free (power law) networks attracted an exceptional amount of attention in the literature. (Newman, 2003) This was mainly motivated by the realization that many real-world networks show this property, including artificially generated ones like the Internet. Since this property apparently contributes to the robustness of these networks, theoretical interest has arisen for models and methods for the generation of scale-free networks.

4.1 Related Work

Despite the high-level of interest, today's models of scale-free networks are mainly variants of the preferential attachment (PA) model of Albert and Barabási (2002). Similarly to our model, the PA model also applies an incremental network generation scheme. Nodes join the network one at a time and create a fixed (*E*) number of links to the existing nodes. However, the selection of nodes to link to, in the PA case, is based on the current degree distribution of the network. More specifically, the probability for a node to be selected is proportional to the number of links it has. This violates the second practical constraint laid down in Section 1 of this paper. Namely, the PA model assumes that the current degree distribution of the network is available to all newcomers at no cost.

Therefore, the second practical constraint that assumes that information about the network is costly, sets our model apart from the work of Albert and Barabási, and from other variants. We believe that this is a rather important difference. The availability of timely information on each node assumes a global point of view and a top-down model. In contrast, the model presented in this paper applies a bottom-up approach, where the network is generated using a limited amount of information only. To our knowledge, this is the first bottom-up model capable of generating robust and scale-free networks.

4.2 Future Work

The results reported in this paper show that a bottom-up optimization process that is constrained by local budget limitations is capable of generating robust networks, including ones with a scale-free degree distribution. However, the particular model presented here is by no means ultimate, nor the work is finished. A number of variants could and should be created that may well result in more efficient or more realistic algorithms that produce more realistic and more robust networks. The pricing schemes studied in this paper are somewhat arbitrary, a number of others have promising potential. Also, it would be worthwhile to apply a more formal measure of robustness to the generated networks, in order to be able to more accurately compare the results of various pricing schemes, and to find the best scheme possible.

One model variant we especially intend to investigate introduces additional economic constraints for the agents. In this model not only information is costly, but so is creating a link. Therefore, agents face the common '*explore or exploit*' decision, and need to trade off links for information.

5. CONCLUSION

In this paper we addressed the problem of generating robust networks that have good chances to survive random failures (i.e., the random removal of nodes). We created an agent-based model to incrementally generate such networks, and explored it by numerical simulation. The key concept of this model is that the network is formed as a result of the individual optimizing actions of agents (nodes) that are subject to economic constraints regarding the agents' information access. The main finding of the simulation experiments is that the model generates robust networks (i.e., ones that are biased towards weekly connected nodes) under a wide variety of parameter settings. Moreover, it is also capable of generating special topologies, like a 'star topology', or a scale-free network.

It is worth noting that the model's economic metaphors (i.e., pricing scheme and budget constraint) are, but a way to define various patterns of information access. The results of the model imply that variations in information access pattern can lead to very different network topologies. On a more abstract level, this observation suggests that control over the information access of a high number of independent actors could turn into a powerful means to control the configuration that emerges at the system level. In other words, conscious design of the patterns of agent-level information access may prove to be valuable as a tool for emergent synthesis.

ACKNOWLEDGEMENT

Part of the computer simulations were done on the SUN E10K, E15K system of Hungary's NIIF Supercomputing Center in Budapest. This support is gratefully acknowledged.

REFERENCES

Albert, R., Barabási, A-L. (2002): Statistical Mechanics of Complex Networks, *Rev. Mod. Phys.* **74**, pp. 47-97.

Cederman, L-E. and Gulyás, L. (2001): Tutorial Sequence for RePast: An Iterated Prisoner's Dilemma Game, http://www.cbrss.harvard.edu/events/cccsn/tutorial.pdf

Erdos, P. and Rényi, A. (1959): On random graphs, *Publicationes Mathematicae* **6**, pp. 290-297.

Faloutsos, M., Faloutsos, P., Faloutsos, C. (1999): On Power-Law Relationships of the Internet Topology, *ACM SIGCOMM Computer Communications Review* **29**(3), pp. 251-262.

Fruchterman, T.M.J., and Reingold, E.M. (1991): Graph Drawing by Force-directed Placement, *Software-Practice and Experience*, **21**(11), pp. 1129—1164.

Newman, M. E. J. (2003): The structure and function of complex networks, *SIAM Review* **45**, pp. 167-256.

RePast, Agent-Based Simulation Framework, (2003) http://repast.sourceforge.net

Watts, D. J. and Strogatz, S. H. (1998): Collective dynamics of 'small-world' networks, *Nature* **393**, pp. 440-442.

---

[i] Networks may have a radically different appearance depending on the layout algorithm used to prepare their visualization. The network figures included in this paper were created using the algorithm of Fruchterman and Reingold (1991), except for the bottom row of Figure 4.

[ii] The source code of the agent-based simulation is available from the author upon request.