An Analysis of Constructive Network Formation Models

by

Gary Fredericks

Bachelor of Science
University of South Carolina, 2010

Submitted in Partial Fulfillment of the Requirements

for the Degree of Master of Science in

Computer Science

College of Engineering and Computing

University of South Carolina

2013

Accepted by:

José Vidal, Major Professor

Duncan Buell, Committee Member

Stephen Fenner, Committee Member

Lacy Ford, Vice Provost and Dean of Graduate Studies

Abstract

We study a family of network formation models to determine how payment rules affect the final network topologies that emerge. In our model a set of nodes starts out without any edges and the nodes must pay for the creation of edges using one of several different payment mechanisms. Example payment mechanisms include one node paying for the whole edge, and the cost being shared equally between the two nodes. We show how the set of networks formed by some payment rules are subsets of those formed by other rules. We also perform extensive empirical tests on networks of up to 10 nodes. These tests reveal some interesting patterns in the connectivity, stability, and fairness of the networks generated by the various payment rules given a fixed link cost.

Table of Contents

Abstract ii

List of Figures iv

Chapter 1 Introduction 1

1.1 Related Work 2

Chapter 2 Model 4

2.1 Payment Rules 6

Chapter 3 Theoretical Analysis 10

3.1 The AE Rule Deadlocks 10

3.2 Relationships Between Payment Rules 12

3.3 Lollipop Graphs 15

Chapter 4 Empirical Tests 19

4.1 Comparing the Sets of Reachable Graphs 20

4.2 Connectivity 22

4.3 Fairness 24

4.4 Regret 26

Chapter 5 Conclusion 30

Bibliography 31

Appendix A The Payment Rule Relation 32

List of Figures

Figure 2.1 Feasibility/Reachability graph diagram 9

Figure 3.1 AE deadlock proof figure 11

Figure 3.2 AE deadlock example 11

Figure 3.3 Partial ordering on rules. 13

Figure 3.4 Partial order incomparability counterexamples 14

Figure 3.5 The order-8 lollipop graph. 15

Figure 3.6 Lollipop graph construction 17

Figure 4.1 Reachable graphs plot 21

Figure 4.2 Expected vertex-connectivity plot 23

Figure 4.3 Expected edge-connectivity plot 24

Figure 4.4 Expected fairness plot 25

Figure 4.5 Expected fairness plot, small α 25

Figure 4.6 Unfairest sink-graph examples 26

Figure 4.7 Regret-free probability plot 28

Figure 4.8 Regret-free probability plot, small α 28

Figure 4.9 Expected regret percentage plot 29

Chapter 1

Introduction

As argued in [2], there is a growing need to understand how real-world networks (such as social networks, supply chain networks, professional networks, the Internet, and overlay networks like BitTorrent) are formed and what characteristics are exhibited by their final topologies. These networks are formed without central control by the collective decision-making of the many selfish agents. We do not currently have the tools to predict what type of network will emerge from a particular set of incentives given to these agents. Specifically, we would like to be able to predict that a given set of incentives will eventually lead to the creation of a network with particular properties, such as high connectivity, high fairness, etc. In this paper we analyze a family of network formation models, each with a different payment rule, and categorize certain characteristics of the network topologies that emerge when these rules are used.

In all the network formation models we study there is a fixed cost for the creation of an edge and the nodes' utility for a topology is proportional to how close they are to everyone else; that is, all nodes want to minimize the path length to all others but do not wish to pay for the new edges. We believe this model (Section 2) captures the essence of the formation of real-world networks without getting into unnecessary details. Future work could extend this basic model and add domain-dependent details to determine if these change the results.

Our goal is to predict the network topology that emerges when we give selfish agents a fixed payment rule. Multiagent systems designers could use our results to choose from different incentive mechanisms depending on what final topological

characteristics they desire. We hope this research will eventually lead to rules such as "if you want the agents to form a network of minimum connectivity d then you should use incentives mechanisms X or Y and set payments as z". The specific characteristics we are currently focusing on are connectivity, fairness, stability, and the variety of distinct topologies that are reachable.

1.1 Related Work

The local connection game, which is what are studying, was introduced in [3], where the authors examine the Nash equilibria of the game and provide some bounds for the price of anarchy, showing that it can be low if the cost of building a new edge is within certain ranges. Specifically, they show that if the cost is less than 2 then the social optimum is the clique and the worst Nash equilibrium is the star, thus the worst price of anarchy is $\frac{3}{4}$. They also show that if the cost of building an edge is greater than 2 then the star is always the social optimum and the price of anarchy is bounded from below by $3 - \epsilon$ and from above by $O(\sqrt{\alpha})$, where α is the cost. They also report having constructed many Nash equilibrium graphs for $\alpha > 2$ and found that they were all trees, except for the Petersen graph, which is a transient equilibrium for $\alpha \leq 4$. However, their Nash equilibrium definition assumes that all vertices choose the set of vertices they want to build at the same time, so the complete graph is built in one step with no intermediate graphs. Our model is incremental.

Corbo and Parkes study the *bilateral connection game* [1] where links must be formed bilaterally. That is, both vertices must agree to build an edge between them and the cost is shared equally between them. They provide upper and lower bounds for the price of anarchy in their model.

Jackson and Wolinsky [5] propose the notion of *pairwise stable equilibrium* in which no vertex wants to drop an adjacent edge and no pair wants to add an edge. In this model vertices can pay for their adjacent edges only. A vertex can unilaterally

fail to pay his share of an edge which results in that edge being dropped. In equilibrium, no vertex wants to stop paying for an edge. The utility each vertex receives is proportional to $\delta^{dist(u,v)}$ where $0 < \delta < 1$, nonlinear in the distance; thus their results cannot be applied to the model used in this paper. Also, in this model vertices are not required to reach all other vertices in the network since there is an "intrinsic value" of communication between every two vertices, which might be 0. A survey of network formation models is given in [4].

Chapter 2

Model

Our model is distinguished from some of the others we referenced by our focus on edge-by-edge construction of the network. We assume that vertices in the network have the goal of minimizing their distance to all other vertices. We also assume that edges have a fixed construction cost, which we refer to as α, and that decisions about whether to build an edge are made based on the immediate improvement in connectivity that the edge will cause in the current network. Thus if vertex b's total distance to the rest of the graph will improve by 5 units by building edge bc, but the cost to b would be 6 units, then b will not want to pay for the edge.

We consider the effect of different rules for who pays, how the cost is divided, and how decisions about whether to build are made. Given such a payment rule, we look at what set of graphs results by starting with an empty network and iteratively adding new feasible edges at random until there are no feasible edges remaining. We do not generally consider the possibility of removing edges that are already built (but see Section 4.4).

We start by defining our basic terms. If N is a positive integer and we let g^N be the set of all subsets of size two of the set $\{1, 2, ..., N\}$, then a **graph** is any $g \subseteq g^N$. That is, a graph is simply the set of all of the edges between its vertices. For the remainder of the paper we will continue to use N to denote the order of the graph, and will denote edges with the short notation ij to represent an edge between vertices i and j.

Given a connected graph g, we let $dist_g(i, j)$ (the **distance** between i and j in g)

be the number of hops in the shortest path from i to j in g (so directly connected vertices always have a distance of 1). If there is no path between two vertices, the distance is infinite. We define a **transition** to be a pair (g, ij) such that $ij \notin g$. So a transition is a graph and an edge that does not already exist (sometimes called an **antiedge**).

Given a transition, we can define the **value** of the new edge for each vertex to be the total increase in connectivity that the vertex obtains with the addition of the edge. That is,

$$v_k = \sum_{l \in N} dist_g(k, l) - dist_{g+ij}(k, l). \tag{2.1}$$

If the edge causes a vertex to be connected to any vertex to which it was not previously connected, then we say that the value of the edge for that vertex is infinite.

A **payment rule** is, informally, a rule for what set of vertices pay for a new edge, how they divide the cost, and how the decision whether or not to invest in the edge is made. Since all vertices behave rationally and will only wish to add an edge if the amount they have to pay is not greater than their value for the edge, we can formally define a payment rule as a predicate $\rho : (g, ij, \alpha) \mapsto \{0, 1\}$. That is, the payment rule determines if a transition can be taken based on α.

Given a payment rule ρ, an edge-cost α, and a transition (g, ij), we say that the transition is **feasible** if the payment rule allows the edge to be added for the cost α. We also use the term **feasibility bound** to denote the maximum α for which a transition is feasible.

We treat **network formation** as an acyclic Markov chain, where the initial state is the empty graph, the other states are all other labeled graphs on N vertices, and a state's outgoing transitions are all feasible transitions to another state, where each has equal probability. So a **network formation process** begins with an empty graph and adds feasible edges at random until there are no more feasible edges to

add.

Given a payment rule ρ and an edge-cost α, we say that a graph g is **reachable** if there is sequence of feasible transitions from the empty graph to the graph g. We can also compute its **reachability**, $r_{\rho,\alpha}(s)$, which is the probability that a network formation process starting with the empty graph is ever in the state represented by g. If a graph is not reachable, then its reachability is 0. Figure 2.1 is a diagram with all of the reachable graphs on 5 vertices and their reachability values for the AP rule with $\alpha = 3$.

A **sink-graph** is any graph that is reachable, but from which no transitions are feasible. Intuitively, this is any graph that can be reached but to which no further edges can be added. The set of sink-graphs for a particular payment rule together with their reachability over α is in some sense a characterisation of the payment rule. Every network formation process is trivially guaranteed to terminate at a sink-graph, so by studying the attributes of sink-graphs we are effectively studying the expected values of the properties of the graphs resulting from a network formation process.

2.1 Payment Rules

We consider six different payment rules that determine which vertices pay for an edge, how the cost is divided, and how the decision is made to build or not.

When deciding which payment rules to study, we considered first that an edge might naturally be paid for by a single adjacent vertex, both adjacent vertices, or the entire graph. Additionally, for the latter two, the cost could be split evenly or in proportion to each vertex's value for the edge. This suggested five payment rules. When we later discovered that the all-vertices-split-evenly payment rule was trivial (Section 3.1), we added a sixth rule with a voting element.

For each of the payment rules we give the expression for the feasibility bound under that payment rule in terms of the values that the vertices have for a transition.

The values of the two vertices incident on the new edge are denoted by v_b and v_c, and v_i ranges over the values of all the vertices in the graph.

Single-Payer (S) This is the traditional game. An edge is created when one of the two adjacent vertices decides to pay for the edge alone. One of the vertices pays α, and the feasibility bound is $\max(v_b, v_c)$. This is the most similar to the model used in [3].

Both-Equally (BE) An edge is created when both vertices agree to split the cost evenly. Both vertices pay $\frac{\alpha}{2}$, and the feasibility bound is $2 \cdot \min(v_b, v_c)$.

Both-Proportionally (BP) An edge is created when both vertices agree to split the cost of the edge in proportion to their values for it. If vertex b and c will reduce their distance to the rest of the graph by v_b and v_c, then b will pay $\alpha \frac{v_b}{v_b+v_c}$ and c will pay $\alpha \frac{v_c}{v_b+v_c}$. The feasibility bound is $v_b + v_c$.

All-Equally (AE) An edge is created when all vertices in the graph agree to split the cost evenly. Each vertex in the order-N graph pays $\frac{\alpha}{N}$, and the feasibility bound is $N \cdot \min_i v_i$.

All-Proportionally (AP) An edge is created when all vertices in the graph agree to split the cost in proportion to their values for it. If the value for vertex i is v_i, then i's payment is $\alpha \frac{v_i}{\sum v_j}$. The feasibility bound is $\sum_i v_i$.

Plurality-Vote (PV) An edge is created when a strict majority of the vertices in the graph are willing to pay their portion of the cost distributed evenly among all vertices. Each vertex in the N-vertex graph pays $\frac{\alpha}{N}$ regardless of their individual votes. We make an exception for transitions which connect disconnected vertices, which we consider to be always feasible. Without this exception, adding an initial edge to an empty graph with $N > 3$ would never be feasible for $\alpha > 0$, as it only benefits two vertices and hence would not be

voted for. The feasibility bound for this payment rule is $N \cdot v_j$ if we assume the v_i are in ascending order and let j be $\lfloor \frac{N-1}{2} \rfloor$.

It is worth noting that the first three payment rules can be thought of as local while the latter three are global.

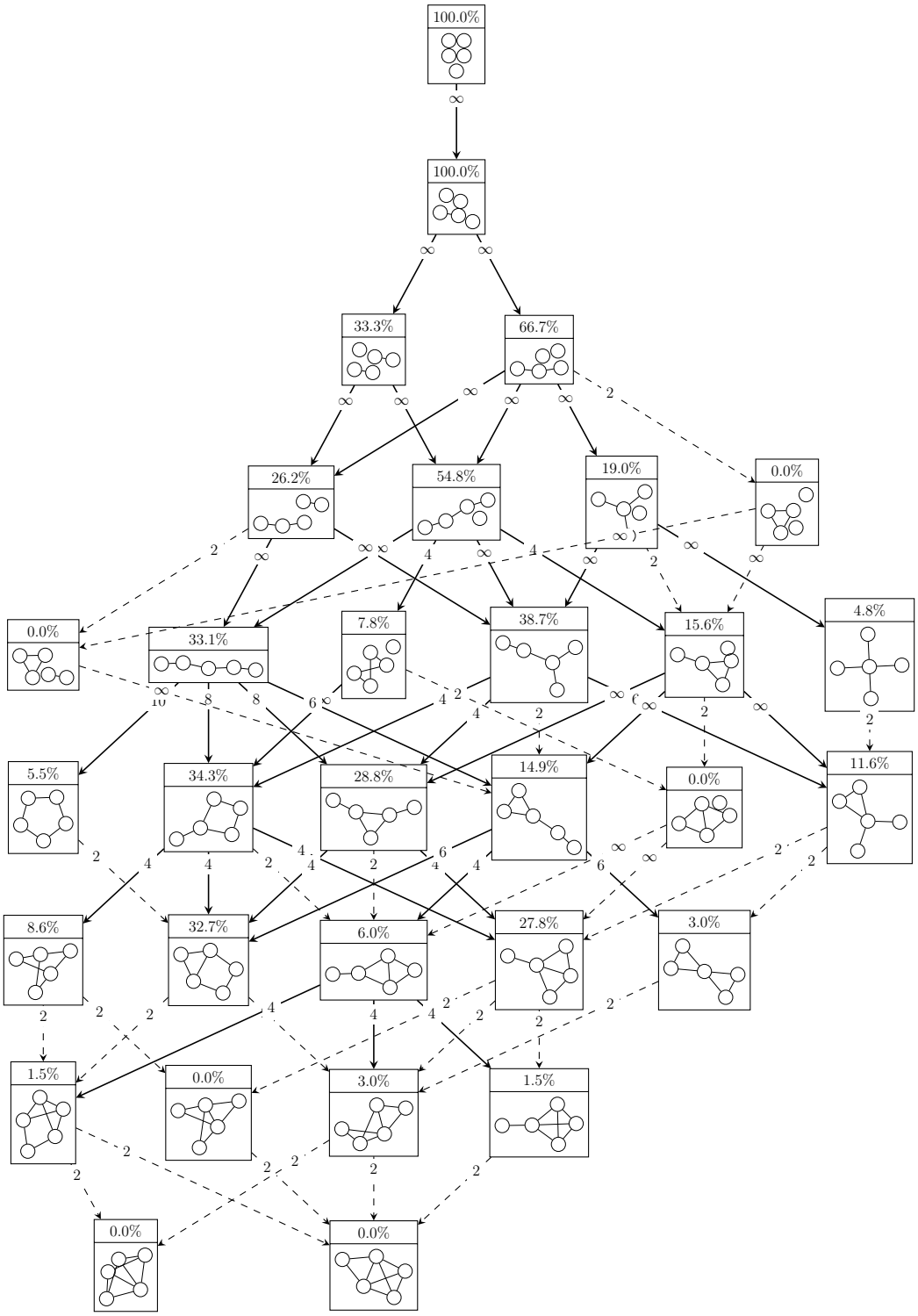

Figure 2.1 An illustration of the relationship between feasibility (the transitions are labeled with their feasibility bounds) and reachability (a percentage marked on the graphs) for graphs of order 5 under the AP payment rule with $\alpha = 3$. Note that only distinct unlabeled graphs are shown, while the reachability values are computed as if summed over all of the underlying labeled graphs.

Chapter 3

Theoretical Analysis

We can prove several interesting facts about these payment rules. We start by showing that the AE rule always deadlocks (and is therefore absent from Section 4). We then define a partial ordering of the six rules under a certain kind of subset relation. Finally, we prove the presence or absence of a certain extreme sink-graph under different rules.

3.1 The AE Rule Deadlocks

Of the six payment rules considered, the AE rule alone has the property of not allowing any feasible transitions out of connected graphs. In the AE rule, for any edge to be added, we require that every vertex in the graph be willing to pay $\frac{\alpha}{N}$ for it. However, for any connected graph and any not-yet-existing edge, there will always be at least one vertex in the graph who is indifferent to the edge, and will not be willing to pay any positive amount for it. The only transitions that are feasible in AE are those that transition from a disconnected graph to a connected graph (which have infinite value for all vertices). This means that a network formation process will always deadlock in the initial state; but even if we made an exception for transitions that connect disconnected parts of the graph (as we do with PV), the process would deadlock as soon as it became a connected tree.

The proof that transitions out of connected graphs are never feasible is fairly straightforward. Consider a connected graph g and two vertices i and j who are not yet directly connected by a link. Since the graph is not disconnected, there is a

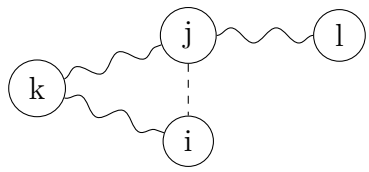

Figure 3.1 Proof that k has value 0 for the edge ij: $dist_{g+ij}(k,l)$ cannot be any shorter by going through ij, because $|\, dist_{g+ij}(k,l) - dist_{g+ij}(k,i)| \leq 1$.

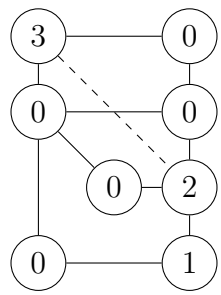

Figure 3.2 An example transition demonstrating the triviality of AE – several nodes have a value of 0 for the transition, in particular the 4 nodes that all lie on midpoints in shortest paths between the two nodes incident on the new edge.

(not necessarily unique) shortest path from i to j. Consider any shortest path as a sequence of vertices, $i, \ldots, j$, of length at least 3. This sequence either has a middle element, if it is of odd-length, or two middle elements, if it is of even length. Consider one of the middle elements in either case, which we can refer to as vertex k. This vertex has the property that $|\, dist_g(i,k) - dist_g(j,k)| \leq 1$.

To show that vertex k is indifferent to the addition of the edge ij, we need to show that the new edge cannot reduce the distance of k to any vertex in the graph (formally, $\forall_{x \in N}\, dist_g(k,x) = dist_{g+ij}(k,x)$). Suppose for the sake of contradiction that there is some vertex l (see Figure 3.1) in the graph for which the shortest path from k to l is shorter in $g+ij$ than in g. This implies that the shortest path in $g+ij$ includes ij. Now suppose without loss of generality that, in the direction from k to l, the path passes first through i and then through j. Then we can show that the path that goes directly from k to j (bypassing i by using the shortest path from k to j in the original graph g) and on to l is at least as short as the path that goes through ij.

The reason is that the distance from k to j directly is at most one unit more than the distance from k to i, and so the distance from k to j through i is at least the same length. Therefore the AE payment rule admits no feasible transitions for $\alpha > 0$. For example transition diagrams (with nodes labeled with their values for the transition) see Figures 3.2 and 3.4.

With the exception of Section 3.2, for the remainder of the paper we will discuss only the other five payment rules.

3.2 Relationships Between Payment Rules

We can construct a partial ordering on the payment rules based on the set of feasible transitions with one rule being a subset of the feasible transitions with another. These relationships are of interest to a system designer because they tell him how his choice of a payment rule will affect the resulting topology. Namely, some payment rules generate possible sets of graphs that are subsets of those generated by other payment rules.

For two payment rules X and Y, we define $X \subseteq Y$ as follows:

$$X \subseteq Y \iff \forall_{g,\alpha,ab \in (g^N - g)} \mathit{Feasible}_X(g, ab, \alpha) \rightarrow \mathit{Feasible}_Y(g, ab, \alpha) \tag{3.1}$$

So $X \subseteq Y$ means that in any network, the feasibility of adding a link with payment rule X implies the feasibility with payment rule Y, which in turn implies that (for a fixed α) the set of reachable graphs with rule X is a subset of (so possibly equal to) the set of reachable graphs with Y. Furthermore we write $X||Y$ for $X \nsubseteq Y \wedge Y \nsubseteq X$.

The following five relations (shown graphically in Figure 3.3) are fairly intuitive, and formal proofs are given in Appendix A.

BE $\subseteq$ BP If both adjacent vertices are willing to split the cost of the edge evenly, they certainly should not mind adjusting the split to be proportional to their

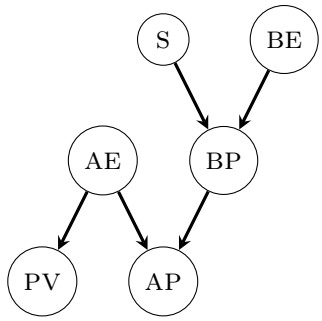

Figure 3.3 Partial ordering on rules.

respective benefits. The fact that the vertex with less to gain is still willing to pay $\frac{\alpha}{2}$ indicates that α is sufficiently low that the vertex with more to gain will be willing to pay its fair share.

S $\subseteq$ BP If either of the two adjacent vertices is willing to shoulder the entire cost of the edge on its own, then splitting the cost proportionally will also be feasible.

BP $\subseteq$ AP If α is low enough for both vertices to split the cost proportionally, then it will be no trouble at all for the rest of the graph to contribute according to their respective values.

AE $\subseteq$ AP A generalization of the (BE $\subseteq$ BP) case, all nodes willing to split evenly always indicates that they would also agree to split proportionally.

AE $\subseteq$ PV Since PV only requires that a strict majority be willing to pay $\frac{\alpha}{N}$, then certainly the case of all vertices being willing is a strict subset.

We have thus established the partial order shown in Figure 3.3. Note that due to the deadlocking of the AE payment rule (Section 3.1), the diagram could also be drawn with AE at the top, pointing at S, BE, and PV.

We can prove by example that all five pairs of rules which are not comparable in Figure 3.3 are indeed incomparable under this relation. That is, we will show that (S || BE) and (PV || AP) and (PV || BE) and (PV || PB) and (PV || S). For a given pair of

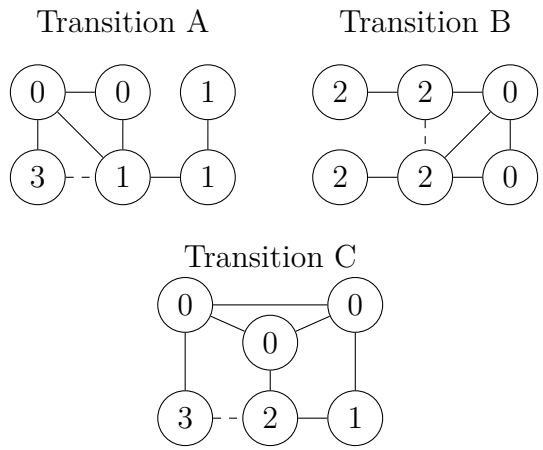

Figure 3.4 Transitions A-C: each vertex is labeled with its value for the new edge, which is drawn dashed.

	PV	AP	BE	BP	S
Transition A	6	6	2	4	3
Transition B	12	8	4	4	2
Transition C	0	6	4	5	3

Table 3.1 The feasibility bounds for the transitions in Figure 3.4. I.e., the maximum α for which the transition shown is feasible under each payment rule.

payment rules, we show this by producing two example transitions: one transition in which feasibility under the first rule implies feasibility under the second, and another transition for which the opposite is true.

Figure 3.4 shows three transitions which suffice to show the incomparability of all five pairs of payment rules. Each graph has an antiedge drawn dashed to indicate the transition. Table 3.1 shows the feasibility bounds (i.e., the maximum α for which the transition is feasible) for each transition and payment rule. The table should be easy to verify by hand.

We can show that BE || S by examining Transitions A and B. In Transition A we have S > BE, while in Transition B we have BE > S. For the other four pairs, which are all of the form PV||$*$, Transition B has PV>$*$ while Transition C has $*$>PV.

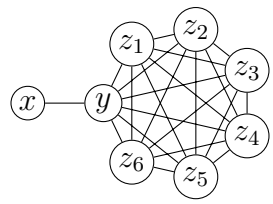

Figure 3.5 The order-8 lollipop graph.

3.3 Lollipop Graphs

For any order N, there is a unique graph (up to isomorphism) for which $N-1$ of its vertices form a clique, and the remaining vertex has only one edge, shared with one of the vertices in the clique (Figure 3.5). This graph has the most edges possible for a graph that can still be disconnected by removing a single vertex. This kind of network would probably be considered undesirable for a variety of applications, and so it is interesting that it appears as a sink-graph in several of the payment rules. We will show that for BE lollipop graphs are never sink graphs, for PV they are never sink-graphs if $N > 5$, but for AP, BP, and S the lollipop graphs are sink graphs for particular ranges of α. We consider lollipop graphs to only exist for $N > 3$, and we will use the name x to refer to the vertex of degree 1, y to refer to the vertex that x is adjacent to, and z_i to refer to the remaining vertices in the clique (as shown in Figure 3.5).

To prove that lollipop graphs are never sink-graphs for the BE payment rule, it will suffice to show that if any of the transitions that lead to the lollipop graph are feasible, then a transition from the lollipop graph is also feasible. Note that there are always exactly three kinds of transitions leading to the lollipop graph (i.e., three kinds of edges that could have been added): edges of the form z_iz_j, yz_i, or xy. In the first two cases, at least one of the vertices would only have value 1 for the edge, because it would not be gaining any connectivity except to the adjacent vertex. For the BE payment rule, this means that the vertices will not agree to add the edge unless

$\alpha \leq 2$. But if $\alpha \leq 2$, any edge at all will be added (since for any transition, both vertices have value at least 1), and so the only sink-graph is the complete graph. This leaves the case of the xy edge being the most recent transition. In this case the state of the network prior to adding the xy edge was a complete connected component of size $N-1$ and the isolated x vertex. So we only have to look one step further back, at the second to last transition which must be either $z_i z_j$ or yz_i, and by the same reasoning as before we must have $\alpha \leq 2$ with the same consequence: the lollipop graph is not a sink-graph.

For the PV payment rule we can prove that lollipop graphs are never sink-graphs for $N > 5$. We can show that for $\alpha > 0$ at most 3 vertices have positive value for last edge, and so it will not be voted for unless either $\alpha = 0$ (in which case the process proceeds all the way to the complete graph) or $N \leq 5$. If the previous edge added is of the form $z_i z_j$, then the two adjacent vertices are the only two who benefit from the edge. If it is of the form yz_i, then only x, y, and z_i have any value for it. Three votes is only a strict majority for $N < 6$, when the lollipop graph is indeed a sink-graph. When the previous edge is xy, we can again look back a second step to the transition that completes the clique; the final edge added to the clique can only benefit the two adjacent vertices, and so will not be voted for when $N > 5$.

For the AP, BP, and S payment rules, we can prove that for certain values of α the lollipop graph will always be a sink-graph. For this we have to prove that for the chosen values of α the lollipop graph is reachable (which requires showing that there is a series of feasible transitions beginning with a tree and ending with the lollipop graph), and that there are no further feasible transations. Luckily we can use the same reachability construction for all three payment rules, so we will describe it generally before proceeding to the details specific to the individual rules.

The construction begins with a linear graph that has x at one end connected to y, and proceeds by constructing cliques of size $(3, 4, 5, ..., N-1)$ (Figure 3.6) at the

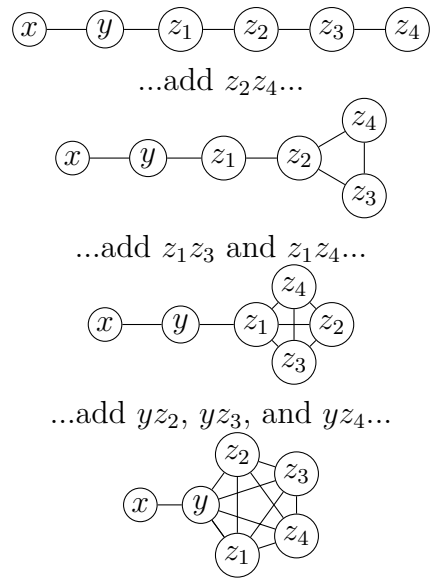

Figure 3.6 Constructing the lollipop graph by incrementally increasing the size of the clique.

end of the line opposite from x. At each step a group of edges (of increasing size) is added, and within the group each edge is indistinguishable and has the same affect on the graph. For the AP, BP, and S payment rules, we need to show that these steps are feasible for a certain α, and that no transitions from the lollipop graph are possible at the same α.

For the AP payment rule, we set $\alpha = 4$. To demonstrate feasibility we just need to show that for any of the edges added in the construction, the total value for the edge over all the vertices in the graph is at least 4. At each step, the new edge connects some z_i to either a z_j or to y. In either case, vertex z_i is reducing its total distance to both y and x, and likewise y and x are reducing their distance to z_i. This means that the total value for the edge over all vertices in the graph is at least 4. Once the lollipop graph is complete, however, any further edge of the form $z_i x$ (which is the only kind of edge possible at that point) can only benefit the two adjacent vertices z_i and x, and so will have a total value of only 2. Therefore within the AP payment rule the lollipop graph will always be a sink-graph for $2 < \alpha \leq 4$.

For the BP payment rule, we set $\alpha = 3$. To demonstrate feasibility for each step we must show that the two vertices adjacent to each edge have a total value for the edge of at least 3. As argued in the preceding paragraph, one of the z vertices adjacent to the edge is reducing its distance to both x and y, and therefore has value at least 2 for the edge. The other adjacent vertex will trivially have value at least 1, and so the total value between the pair is at least 3. Once the lollipop graph is complete, neither vertex has value greater than 1 for any new edge, and so the total value between the pair of vertices will only be 2. Therefore within the BP payment rule the lollipop graph will always be a sink graph for $2 < \alpha \leq 3$.

The S payment rule with $\alpha = 2$ follows very quickly from the argument about BP, because all we need is to note that at least one of the two vertices adjacent to the new edge always has value at least 2 for it, and once the lollipop graph is constructed no vertex has value greater than 1 for any new edge. Therefore within the S payment rule the lollipop graph will always be a sink-graph for $1 < \alpha \leq 2$.

These results for lollipop graphs are especially interesting when considering the implications for modified models allowing new nodes to join existing networks. Given the right payment rule and conditions, a new node joining an already-complete network who creates a single link to one of the existing nodes may be economically unable to add links to any other nodes, leaving it at much higher risk of disconnection than the rest of the network.

Chapter 4

Empirical Tests

We also investigate the payment rules experimentally by exhaustively computing properties of all connected graphs on ten or fewer vertices, and try to identify patterns that seem most amenable to extrapolation to larger graphs. Most of our experimentation was done by treating the network formation process as an acyclic Markov chain, where the Markov states are graphs, the transitions represent adding an edge to the graph, and the transition probabilities are conditioned on α. We started by creating a database with all unlabeled graphs for orders three through ten, for a total of 12,293,431 graphs. We also generated all possible transitions between the graphs, which totaled 251,463,867. Data generation was aided by the `nauty` program [6]. We used unlabeled graphs because tracking labeled graphs explicitly would be infeasible (there are over 35 trillion of them for $N \leq 10$), and it was possible to perform the calculations so as to obtain the same result as we would have had for labeled graphs.

The next step toward analyzing the behavior of network formation processes was to compute the feasibility thresholds for each payment rule and transition. I.e., given a payment rule and transition, what is the largest value of α for which the transition is feasible? Using these feasibility thresholds, we computed the reachability of each graph and payment rule. To calculate the reachability $r_{\rho,\alpha}(s)$ of each graph s we use dynamic programming by first setting the r-value of the empty graph (the one with no edges) to 1. Then for each successive graph size (from 1 to $\frac{N(N-1)}{2}$) we compute the reachability of each graph g based on the reachabilities of the graphs that transition to it:

$$r_{\rho,\alpha}(g) := \sum_{g' \in \text{neighbors}(g)} r_{\rho,\alpha}(g') \cdot p_{g'g} \tag{4.1}$$

where neighbors(g) is the set of all graphs that can transition to g, $p_{g'g}$ is the probability that the graph g' will transition into g by the addition of one edge, and ρ is a payment rule. These reachability values give us the probability that a particular graph will be reached given that we started with the empty graph. For those graphs s that are sink graphs—they do not have any outgoing transitions, given α and ρ—the reachability value $r(s)$ tells us the probability that sink graph s will be the one reached by the network formation process.

The rest of this section describes some of our methods for comparing the payment rules using the generated data. We start by comparing the number of reachable graphs for each payment rule as α varies. Then we compute and compare two different attributes of the sink-graphs, which are connectivity and fairness. Finally we look at the proportion of the sink graphs which have a certain equilibrium property.

4.1 Comparing the Sets of Reachable Graphs

We first examine the number of reachable graphs for a given payment rule as α changes. The number of reachable graphs changes as a step function over α for all our payment rules. We can say that the points at which the function changes are the **critical $\boldsymbol{\alpha}$ values** for that payment rule and order. These critical α values are always integers for all of our payment rules because, for each rule, a transition's feasibility bound is determined by a function of the vertices' values. These functions are integer valued when their inputs are integers, and the vertices' values are always integral.

There are further restrictions on the critical α values for three of the payment rules. For the PV rule the critical α values are always multiples of N. This is because the feasibility bounds occur when $\frac{\alpha}{N}$ is not greater than a majority of the vertices' values, which are all integers.

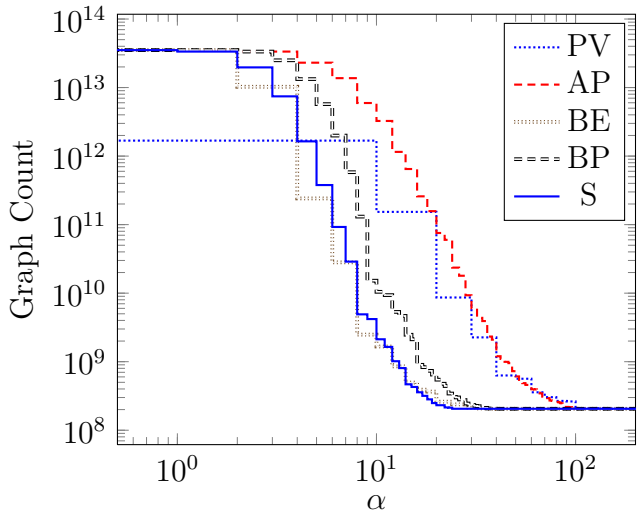

Figure 4.1 Number of reachable order-10 graphs for each payment rule as α changes. Since the x-axis is plotted logarithmically, the lines extend infinitely in both directions. At $\alpha = 0$ all lines meet at the same value.

For the AP and BE rules, the critical α values are always multiples of 2. For AP this is because all vertices' values are considered (the feasibility bound is simply the sum of all values), and for a particular transition the values will be symmetric: for any vertex a who has positive value for the new edge because its path to b has been shortened, b's path to a will have been shortened as well, and so the value is counted twice. For the BE rule, the critical α values are always even for the same reason that the critical α values for PV are always multiples of N: the feasibility bounds occur when $\frac{\alpha}{2}$ is not greater than either of the two adjacent vertice's values, which are both integers.

Figure 4.1 shows a plot of reachable graphs for order 10 with both axes logarithmic. Note that, of the five payment rules, PV is the only one for which some graphs are not reachable for any $\alpha > 0$. This is why the PV line is below the others as α goes to zero.

The theoretical results in Section 3.2 are evident in Figure 4.1. The line for AP dominates the BP line, which itself dominates both S and BE. These facts correspond to the subset relationship described in Section 3.2 . The proofs of incomparability

correspond to the facts that the S and BE lines do not have a dominance relationship, and that the PV line has no dominance relationship with any other line.

We also notice in Figure 4.1 that the curve for PV starts with a lower value than all other curves and this value remains constant even as the other curves decrease, such that by the time $\alpha = 10$ all the other payment rules, except for AP, have a lower value. Our test results for all other orders less than 10 show a similar pattern: the number of graphs for all other payment rules start out higher than for PV but they all drop below PV by the time $\alpha = N$. We hypothesize that this pattern will hold for all orders, namely, that the number of graphs produced by PV will always be lower than BE, BP, and S for small α but this relationship will be the reverse by the time $\alpha = N$.

4.2 Connectivity

Our second method of comparing payment rules is measuring the connectivity of the sink-graphs that they produce. Connectivity has two formalisms: vertex connectivity is the minimum number of vertices that can be removed before the remainder of the graph is disconnected, while edge connectivity is the minimum number of edges with the same effect. The two types of connectivity measure the stability of the network against vertex failures and edge failures respectively. In both cases a connectivity of 0 is synonymous with a disconnected graph, while a connectivity of $N - 1$ (for order N) is a fully-connected graph.

Vertex Connectivity

Figure 4.2 shows the expected vertex-connectivity of the sink graphs as α varies for the five payment rules and order 10. These expected values are simply the summation over the product of the probability that a particular graph exists (its reachability)

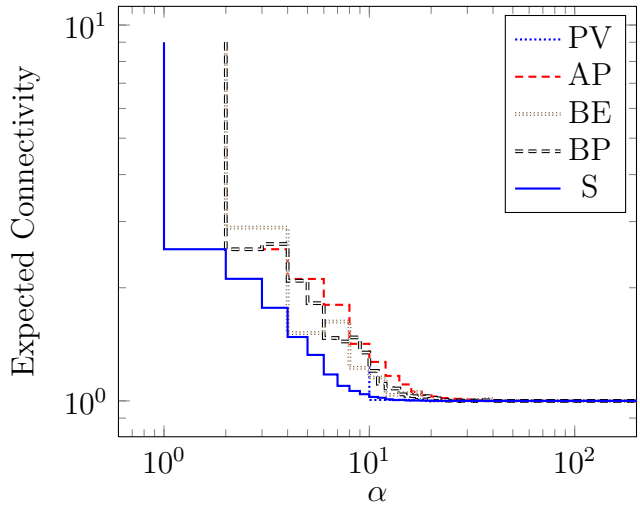

Figure 4.2 Expected vertex-connectivity of sink-graphs of order 10 by α.

and its connectivity. Thus, the values tell us what connectivity the designer of a system can expect given a chosen payment rule and α value.

Figure 4.2 shows that, as expected, connectivity falls as α increases, but it also shows that it falls faster for S than for all the other payment rules. In fact, S is almost dominated by all other rules, and all the other rules are non-comparable: they intersect each other for some α. Thus, if vertex-connectivity is a desired feature of the final network then one should avoid the single payer payment rule. We also note that despite what a glance at the figure suggests, none of the payment rules have connectivity decreasing monotonically, an unexpected result.

Edge Connectivity

For most graphs, the vertex connectivity and the edge connectivity are the same, so the results here are similar. In fact Figure 4.2 is hardly distinguishable from Figure 4.3.

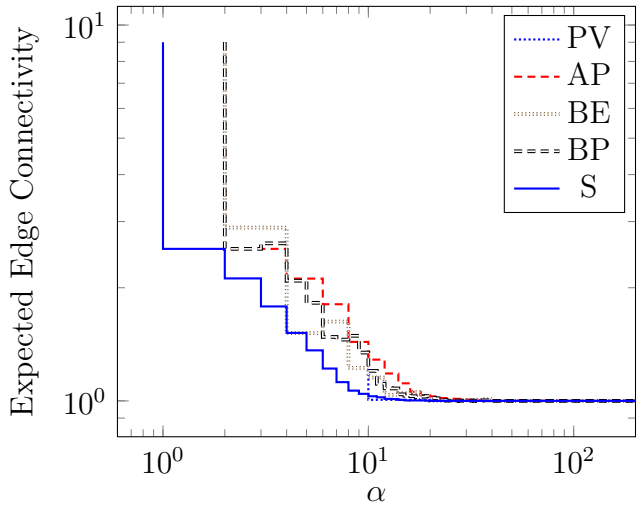

Figure 4.3 Expected edge-connectivity of sink-graphs of order 10 by α.

4.3 Fairness

A third method of comparison is the fairness of the sink-graphs generated by the payment rules. We measure fairness as the difference in connectivity distance between the most connected vertex and the least connected vertex, thus a fair solution can be considered egalitarian. Formally if d_i is the sum of vertex i's distance to all other vertices, then

$$\text{Fairness}(g) = \max_{i \in g}(d_i) - \min_{i \in g}(d_i). \tag{4.2}$$

By this definition a graph with a higher fairness value is less "fair" in the normal sense of the word.

Figure 4.4 shows the expected fairness of the sink-graphs of each payment rule plotted by order. The S rule clearly has the worst fairness of all the rules, while AP and PV are the most fair, except $\alpha < 4$ where BE is most fair, as can be seen in Figure 4.5. Thus, as the cost rises the payment rules that involve all the agents (PV and AP) generate more egalitarian graphs, in terms of connectivity for each node, but when costs are low it suffices to have payment rules that only involve those directly

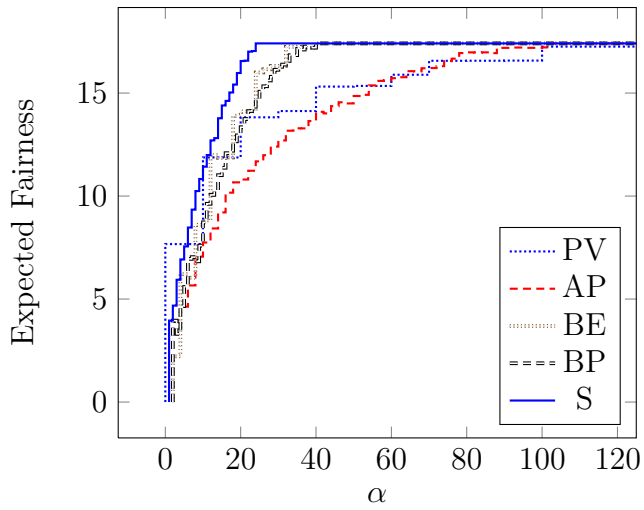

Figure 4.4 Expected fairness for order 10 by α. Lower values are more "fair" in the conventional sense.

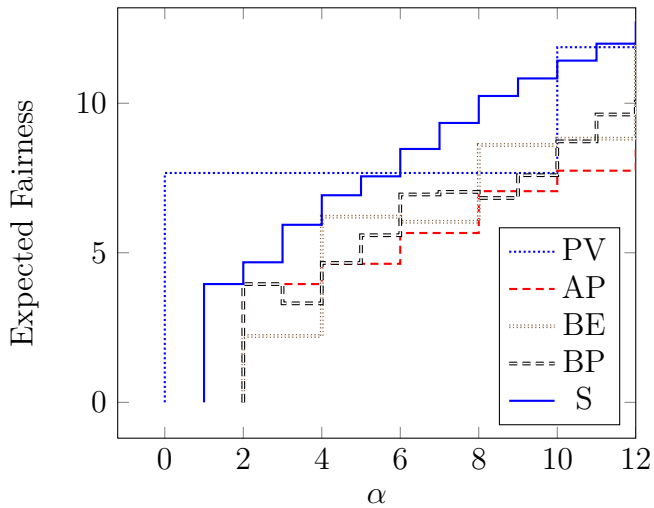

Figure 4.5 Expected fairness for order 10 for small α.

concerned (BE and BP).

A second interesting point is the extent of unfairness reached by all five payment rules. BE and BP are a bit milder for only reaching 18, while the other three reach 20. Figure 4.6 shows examples of the unfairest graphs for each payment rule.

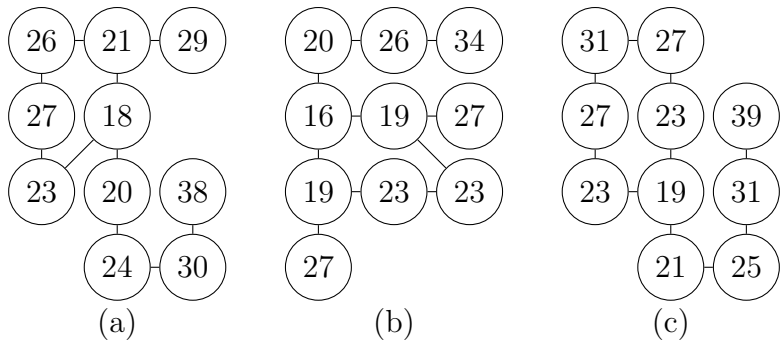

Figure 4.6 Examples of the unfairest sink-graphs of each payment rule for $N = 10$: Graph (a) has *Fairness* $= 20$ and is a sink-graph for AP and PV. Graph (b) has *Fairness* $= 18$ and is a sink-graph for BE and BP. Graph (c) has *Fairness* $= 20$ and is a sink-graph for S.

4.4 Regret

Regret-Free Sink Graphs

A fourth comparison method is a stronger equilibrium concept than sink-graphs, where we consider not only the ability to add further edges, but also to remove existing edges. More accurately, we do not consider the effects of actually removing edges (which could change our definition of reachability), but we do note for a sink-graph whether its existing edges are worth adding back if they were removed. For this reason the most descriptive term seems to be **regret**, and so in this section we investigate the probability that the sink-graph reached by a network formation process is **regret-free**. Formally, g is regret-free if

$$\forall_\alpha(Reachable(g) \rightarrow \forall_{e \in g} Feasible(g - e, e)). \tag{4.3}$$

Figure 4.7 shows the probabilities that a sink-graph is regret-free, for order 10. The probabilities start at 1 for $\alpha = 0$ because for low-enough α the only sink-graph is the complete graph, which is always trivially regret-free. Similarly the probabilities climb slowly back up to 1 as α increases because more and more of the sink-graphs are trees, which are also trivially regret-free. For small values of α being regret-free

is relatively rare (see Figure 4.8 for the same chart limited to small α). In fact, if we examine the global minimum of the probability functions as a function of the order, it decreases drastically toward 0 as the order goes up to 10. For $N = 10$, the minimum probabilities are roughly 10^{-6} for S, BP, and AP, 0.0047 PV, and 0.012 for BE. All payment rules have their minimum probabilities approaching 0, though notably PV and BE are approaching slower. It is also interesting that the minimum probabilities are always positive up to order 10, when there is no obvious reason they could not be 0. Despite the differences, the extremely low percentages probably indicate that the metric may not be very relevant in practice.

Despite their relative rarity, the regret-free sink-graphs we computed have an interesting property: they tend to be disproportionately symmetric. That is, there is a non-trivial way to permute the vertices such that the same graph results. In fact, for all the graphs we examined (up to $N = 10$) we were not able to find any asymmetric regret-free sink-graphs with the exception of trees. This is surprising because for sink-graphs generally, the proportion of asymmetric graphs is quite substantial and rises with N: for $N = 8$, at least 34% of sink-graphs are asymmetric for all payment rules, and 54% and 61% for $N = 9$ and $N = 10$ respectively. We thus hypothesize that regret-free non-tree sink-graphs are symmetric for all orders.

Regret Percentage

Since being purely-regret free seems to be rare enough as to be considered practically unachievable, it is natural to generalize the property by measuring the percentage of the edges of the sink-graph that are "regretful". If we measure the expected percentage of regretful edges, we get the chart in Figure 4.9.

The y-axis is somewhat inverted compared to the previous chart, since the higher positions signify more regret, while in the previous chart they signified probability of being regret-free. The lines are interestingly erratic, but we can see some of the

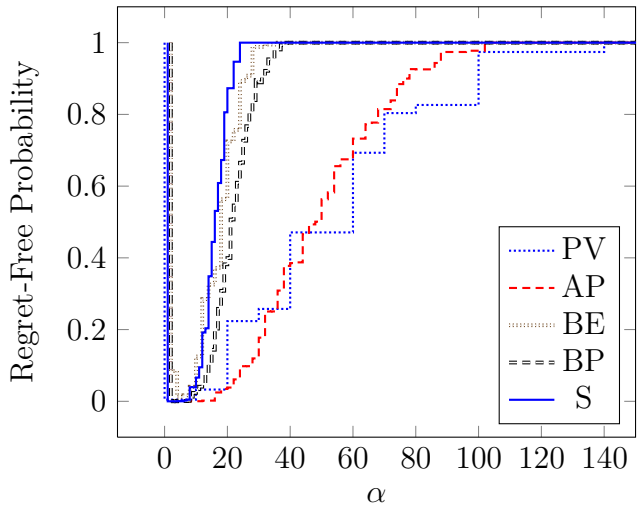

Figure 4.7 Probability that a sink graph in an order-10 network formation process is regret-free.

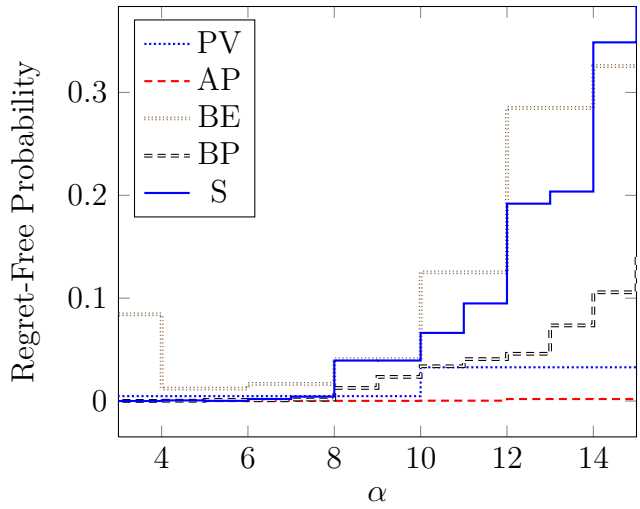

Figure 4.8 Probability that a sink graph in an order-10 network formation process is regret-free, for small α.

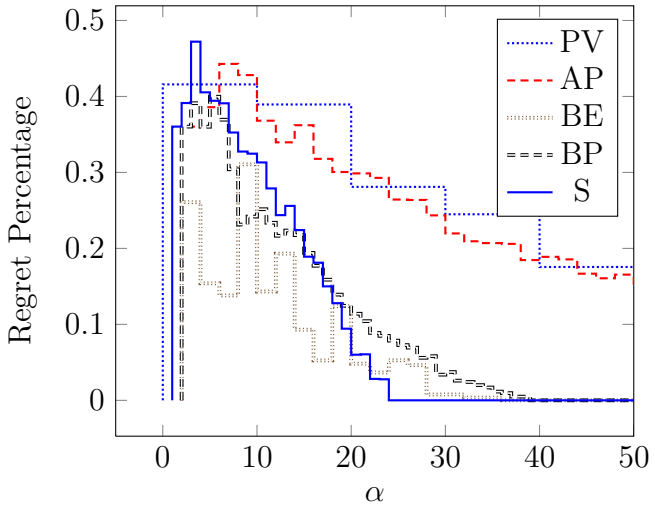

Figure 4.9 Expected regret percentage for sink-graphs in an order-10 network formation process.

same trends between payment rules as in the regret-free-probability measurement, with the global rules (PV and AP) doing markedly worse than the local ones. In this case however we get more information about what happens in the average case: a significant number of edges are likely to be regretted, but generally less than half.

Chapter 5

Conclusion

Our work lays the foundation for interesting research into ways to predict the topology of graphs that emerge from network formation processes that rely on selfish agents. We studied particular payment rules and were able to show that the AE rule has no feasible transitions from connected-graphs for $\alpha > 0$. We were also able to prove the partial order shown in Figure 3.3, which relates directly to the number of reachable graphs for the payment rules, and correlates with the average connectivity of the sink-graphs. Additionally we were able to prove that a certain rather extreme graph, the lollipop graph, is certain to be a sink-graph for AP, BP, and S, but virtually never for BE and PV.

Our empirical research showcases a number of regular patterns: the average sink-graph connectivity is lowest for S, the AP and PV rules seem more fair than BE and BP for high costs but vice-versa for low costs, and regret-free sink-graphs seem to be always symmetric. The consistency of the patterns as N increases to 10 offer strong evidence that they will continue for larger N. Our ongoing work deals with expanding our formal analysis to explain the patterns found by our empirical work.

There are a number of areas that we touched that we think deserve deeper study. For example, we have assumed that network formation only moves forward (adding edges) and never backward (removing edges), which is likely an unacceptable assumption for many applications. There is also much theoretical and experimental work that could be done to study the accuracy of the predictions we made regarding larger graphs.

Bibliography

[1] Jacomo Corbo and David Parkes. The price of selfish behavior in bilateral network formation. In *Proceedings of the twenty-fourth annual ACM symposium on Principles of distributed computing*, pages 99–107, New York, NY, USA, 2005. ACM.

[2] David Easley and Jon Kleinberg. *Networks, Crowds, and Markets: Reasoning About a Highly Connected World*. Cambridge University Press, 2010.

[3] Alex Fabrikant, Ankur Luthra, Elitza Maneva, Christos H. Papadimitriou, and Scott Shenker. On a network creation game. In *Proceedings of the twenty-second annual ACM symposium on Principles of distributed computing*, pages 347–351, New York, NY, USA, 2003. ACM.

[4] Matthew O. Jackson. A survey of models of network formation: Stability and efficiency. In G. Demange and M. Wooders, editors, *Group Formation Economics: Networks, Clubs, and Coalitions*. Cambridge University Press, 2005.

[5] Matthew O. Jackson and Asher Wolinsky. A strategic model of social and economic networks. *Journal of Economic Theory*, 71(1):44–74, 1996.

[6] Brendan D. McKay. *nauty User's Guide (Version 2.4)*. `http://cs.anu.edu.au/~bdm/nauty/`.

Appendix A

The Payment Rule Relation

Here we present formal proofs of the relationships listed in Section 3.2.

BE $\subseteq$ BP $\wedge$ AE $\subseteq$ AP We can first prove AE $\subseteq$ AP and then show that the proof extends to BE $\subseteq$ BP. Start by assuming that a transition is feasible under the AE payment rule. This means that every vertex is willing to pay $\frac{\alpha}{N}$ for the edge, which means that $\forall_{i \in N} v_i \geq \frac{\alpha}{N}$. Under the AP payment rule each vertex is require to pay $\alpha \frac{v_i}{\sum v_j}$, and so a transition is only feasible when $\forall_{i \in N} v_i \geq \alpha \frac{v_i}{\sum v_j}$. The expression can be simplified to $\alpha \leq \sum v_j$, and this follows with a bit of arithmetic from the earlier assumption that $\forall_{i \in N} v_i \geq \frac{\alpha}{N}$. The proof that BE $\subseteq$ BP is similar, but deals with just two vertices instead of the whole graph.

S $\subseteq$ BP $\wedge$ BP $\subseteq$ AP The claims that S $\subseteq$ BP and BP $\subseteq$ AP can be generalized to the claim that if a group of vertices agree to pay for an edge with the cost divided proportionally among them, any superset of that group will also want to build the edge (this requires recognizing that the S payment rule could be described as trivially dividing the payment proportionally among one vertex). Consider any two non-empty sets of vertices B and C such that $B \subseteq C$. Assume that the vertices in B are willing to split the cost of an edge proportionally, from which we can conclude that $\forall_{i \in B} v_i \geq \alpha \frac{v_i}{\sum v_j}$. This expression is not dependent on v_i as it simplifies to $\alpha \leq \sum_{j \in B} v_j$. To show that all vertices in C will also agree to split the cost of the edge proportionally we need to derive the same expression for C, namely that $\alpha \leq \sum_{j \in C} v_j$. This is not hard, since clearly $\sum_{j \in B} v_j \leq \sum_{j \in C} v_j$, which gives us $\alpha \leq \sum_{j \in B} v_j \leq \sum_{j \in C} v_j$, which is sufficient.

AE $\subseteq$ PV This is the easiest relationship to prove, because AE and PV both require payments of $\frac{\alpha}{N}$ from each vertex. The only difference is the decision of whether or not to add the edge, which requires unanimity in the case of AE, but only a strict majority in the case of PV. Clearly the former is a subset of the latter.