A Clustering Coefficient Network Formation Game

Michael Brautbar Michael Kearns
brautbar@cis.upenn.edu , mkearns@cis.upenn.edu
Department of Computer and Information Science
University of Pennsylvania

Abstract

Social and other networks have been shown empirically to exhibit high edge clustering — that is, the density of local neighborhoods, as measured by the clustering coefficient, is often much larger than the overall edge density of the network. In social networks, a desire for tight-knit circles of friendships — the colloquial “social clique” — is often cited as the primary driver of such structure.

We introduce and analyze a new network formation game in which rational players must balance edge purchases with a desire to maximize their own clustering coefficient. Our results include the following:

- Construction of a number of specific families of equilibrium networks, including ones showing that equilibria can have rather general tree-like structure, including highly asymmetric trees. This is in contrast to other network formation games that yield only symmetric equilibrium networks. Our equilibria also include ones with large or small diameter, and ones with wide variance of degrees.
- A general characterization of (non-degenerate) equilibrium networks, showing that such networks are always sparse and paid for by low-degree vertices, whereas high-degree “free riders” always have low utility.
- A proof that for edge cost $\alpha \geq 1/2$ the Price of Anarchy grows linearly with the population size n while for edge cost α less than 1/2, the Price of Anarchy of the formation game is bounded by a constant depending only on α, and independent of n. Moreover, an explicit upper bound is constructed when the edge cost is a ”simple” rational (small numerator) less than 1/2.
- A proof that for edge cost α less than 1/2 the average vertex clustering coefficient grows at least as fast as a function depending only on α, while the overall edge density goes to zero in a rate inversely proportional to the number of vertices in the network.
- Results establishing the intractability of even weakly approximating best response computations.

Several of our results hold even for weaker notions of equilibrium, such as those based on link stability. We also consider other variants of the game, including a non-normalized version of clustering coefficient and bilateral edge purchases one.

1 Introduction

The proliferation of large-scale social and technological networks over the last decade has given rise to an emerging science. One of the primary aims of the empirical branch of this new science is to quantify and examine the striking apparent structural commonalities that many of these large networks share, despite their differing origins, populations, and function. For example, one empirical narrative in this vein that is still unfolding is the claim that large-scale networks from social, economic, technological, biological and other origins often share the properties of small diameter, heavy-tailed degree distributions, and high edge clustering.

Because of this, one of the primary goals of the theoretical branch of this new science is the formulation of simple models of network formation that can explain such apparent structural universalities. Interestingly, to date such efforts have mainly fallen into two categories. In the stochastic network formation literature, probabilistic models for network growth are proposed that exhibit one or more of the structural universals of interest in expectation or with high probability. In contrast, in the game-theoretic network formation links do not form randomly, but for a "reason" (rationality), and the interest is in the structural and other properties that can arise at population equilibrium. The game-theoretic models to date have primarily technological, rather than sociological, motivations, such as efficient routing concerns in communication networks (see [18, 12] for good overviews of both approaches, as well as Related Work below).

In this paper we introduce and study a new network formation game explicitly motivated by an empirical phenomenon often cited in large social networks: the tendency for friendship to be transitive, or for friends of friends to be friends themselves [12, 7]. In sociology and other fields, this notion is quantified by the *clustering coefficient* of a network, and a long series of studies has documented the fact that social networks routinely exhibit much larger clustering coefficients than would be expected from their overall edge density alone [19, 12]. In social networks, homophily (the tendency for like to associate with like), the tendency for introductions to be made through mutual acquaintances, and a human desire for tight-knit cohorts are all cited as possible forces towards high clustering coefficients [11, 7]. Given the frequent observation of clustering in social networks, and the long history of sociological and psychological theories regarding its origins in individuals, it is of interest to examine the consequences when clustering is considered the primary source of utility in a network formation game. In the same way that previous papers have taken abstract human or organizational desires, such as those of being well-connected or centrally placed in a network, and studied them as network formation games [13, 3, 10, 9], here we do so for the notion of clustering.

We thus introduce and analyze a network formation game in which rational players must balance edge purchases, each of fixed cost, with a desire to maximize their own clustering coefficients. Like most of the prior work in formation games, we consider a unilateral, rather than bilateral, edge purchase model (Twitter rather than Facebook); such a model is appropriate for many, though obviously not all, social networks. Our results include the following:

- Construction of a number of specific families of equilibrium networks, including ones showing that equilibria can have rather general tree-like structure, including highly asymmetric trees. This is in contrast to other network formation games that yield only symmetric equilibrium networks. Our equilibria also include ones with large or small diameter, and ones with wide variance of degrees.

- A general characterization of (non-degenerate) equilibrium networks, showing that such networks are always sparse and paid for by low-degree vertices, whereas high-degree "free riders" always have low utility.

- A proof that for edge cost $\alpha \geq 1/2$ the Price of Anarchy grows linearly with the population size n while for edge cost α less than $1/2$, the Price of Anarchy of the formation game is bounded by a constant depending only on α, and independent of n. Moreover, an explicit upper bound is constructed when the edge cost is a "simple" rational (small numerator) less than $1/2$.

- A proof that for edge cost α less than $1/2$ the average vertex clustering coefficient grows at least as fast as a function depending only on α, while the overall edge density goes to zero in a rate inversely proportional to the number of vertices in the network.

- Results establishing the intractability of even weakly approximating best response computations.

Several of our results hold even for weaker notions of equilibrium, such as those based on link stability. We also consider other variants of the game, including a non-normalized version of clustering coefficient and bilateral edge purchases one.

2 Related Work

Models of social and technological networks can be roughly divided into two categories — stochastic generative models and game-theoretic models.

A stochastic generative model captures the dynamics of a specific stochastic process and characterizes the networks created in the limit of that process. Perhaps the most notable stochastic generative models are the preferential attachment model [4] and the small-world model [20]. In the preferential attachment model nodes arrive one at a time and each new node stochastically connects to a fixed number of previous nodes, where the probability of connecting to a specific node is proportional to that node's current degree in the network. Networks created by the model are known to have a limiting power-law degree distribution [5], a prominent property of various social networks. In contrast to the preferential attachment model, the small-world generative model assumes that all nodes are given in advance. In that model one starts with a ring lattice on the n nodes and rewires each edge independently with some fixed probability. Networks created this way are known to have low diameter and a large average clustering coefficient, for a large range of the rewiring probability [20]. While the preferential attachment model and the small-world model are able to only generate networks with some properties of real social networks, a recent model following similar lines as that of preferential attachment was shown to being able to generate networks with several more properties of real social networks [17].

A second approach to modeling social and technological networks is based on game theory. A node is equipped with a utility function that for each outcome of the game quantifies how good the outcome is for that node. The utility of a node is a function that depends on the structure of the outcome network and the cost of the edges the node purchased. Game theoretic formation models roughly divide into unilateral and bilateral games. In unilateral games a node can purchase an edge to another node without asking for that node's consent. In a bilateral game each edge is a result of mutual consent between the edge's endpoints. In both variants once the edge is constructed both parties can use it[1]. The seminal work of Fabrikant et al. [10] present an Internet routing latency game where the utility of a node is the sum of its shortest path distances to all other nodes plus the cost of the edges the node purchased. The game is perceived as a minimum latency game where a node's goal is to route packets quickly to their destination. The authors showed that regular trees are Nash Equilibrium (NE) networks of the game and raised the question whether the game has non-tree NE. Albers et al. [1] provide the first construction of a cyclic NE for the game using methods from finite affine spaces. Alon et al. [2] provide a combinatorial construction of a link stable network with diameter three for the routing game.

Bala and Goyal analyzed a general formation game where the utility of a node is a two-parameter function where the first parameter is the number of nodes a node is connected to in the outcome graph, and the second parameter is the number of edges the node bought [3]. Under a mild monotonicity condition on this utility function the authors showed that the Nash Equilibrium networks of the game are trees and the strict Nash Equilibrium networks are star-like (plus the empty network for some edge costs).

Borgs et al. [6] have recently introduced a unilateral network formation game motivated from affiliation networks. In their model a player can unilaterally initiate social events with a cost proportional to the number of invitees. Any two players that meet regularly at events will then form an (undirected) edge. The utility of a player is its degree in the network minus the cost of events he initiated. The authors show that the class of NE of the game contains sparse networks as well as power-law networks and that the average clustering coefficient of each NE network is at least as big as the inverse of the average degree in that network.

Jackson and Wolinsky [13] were the first to introduce a general bilateral game, called the "connections model". The utility of a node in this game is a sum of discounted shortest path distances to all other nodes plus the cost of the edges adjacent to the node. The authors presented the notion of link-stability where no two nodes want to purchase a missing edge between them, and no node wants to unilaterally remove an

[1]A third type of formation games, where edges are purchased unilaterally and can only be used by the purchasing party, is rarely considered in the literature.

adjacent edge. The authors presented a partial characterization of all link stable networks of the game. A specific version of the game, where the edge cost is not uniform but depends on a metric on the network nodes, was further analyzed in [14]. The authors showed that for a specifically chosen discount factor for the utility of path lengths the link-stable networks of the metric version include regular networks, complete networks, chain, and star networks. However, the analysis is limited to specific values of the discount factor and no general characterization of equilibria networks is given.

Inspired by a stochastic model of Kleinberg [15], Kearns and Even-dar introduce a network formation game where players must purchase edges at distance d with cost d^α and wish to minimize the sum of edge purchases and their average distance to other players [9]. The authors show an interesting threshold phenomena for the diameter of the network: The diameter of any Nash Equilibria network is constant for $\alpha < 2$ while for $\alpha > 2$ the diameter of any Nash Equilibria network grows as the square root of the network size.

Evan-Dar et al. [8] analyze a network formation game for bipartite exchange economies. The network is bipartite containing buyers on one side and sellers on the other and edge purchases represent trading opportunity between its endpoint parties. The authors where able to provide a complete characterization of all NE of the formation game, which is rather exceptional in the network formation game literature.

The Price of Anarchy measure was introduced by [16] to quantify the inefficiency of NE networks with respect to a central designed solution. It is defined as the best welfare (sum of node's utilities) of a network to the worst welfare of a NE network. The routing game presented by Fabrikant et al. was shown to have a low Price of Anarchy [10, 1].

3 Preliminaries

The game we shall study, which we will refer to as the *CC game*, is a one-shot, full information game on n players that shall form the vertices of an undirected graph or network. The pure strategies of the game are the possible sets of undirected edges a player may purchase to the other $n-1$ players. The price of all edges is the same and is known in advance to all players. The edge price is denoted by α.

As in a number of previously studied network formation games, we consider edge purchases to be *unilateral* — a player may purchase an edge to any other party without consent from that party — but all players may potentially benefit from the edge purchases of others. In this sense edges are undirected, but we also need to keep track of who purchased each edge. Given the edge purchases of all players the outcome of the game yields a directed network on n nodes, denoted as G , where an edge from node u to node v is present iff u purchased an edge to v. Throughout we shall analyze both the properties of the directed graph G , as well as the undirected graph it induces.

We denote by I_v the set of nodes that purchased edges to v and by O_v the set of nodes v purchased an edge to. We denote the in-degree of v in G as $\textit{in-deg}(v)$ and its out-degree as $\textit{out-deg}(v)$. The total degree of v is defined as $\textit{deg}(v) = \textit{in-deg}(v) + \textit{out-deg}(v)$.

We denote the number of triangles that v is part of in G by $\Delta(v)$. The number of triangles containing v in which the two other nodes both belong to I_v is denoted as $\Delta_I(v)$. Similarly, the number of triangles containing v where the two other nodes belong to O_v is denoted as $\Delta_O(v)$. The number of triangles containing v where one of the other nodes belongs to I_v and the remaining one belongs to O_v is denoted as $\Delta_{I,O}(v)$. These sets are all disjoint by definition and we have $\Delta(v) = \Delta_I(v) + \Delta_O(v) + \Delta_{I,O}(v)$.

The *clustering coefficient* of a node v in G is defined as the probability that two randomly selected neighbors of v are directly connected to each other: $CC(v) = \frac{\Delta(v)}{\binom{deg(v)}{2}}$ if $\textit{deg}(v) \geq 2$, and 0 otherwise. In the CC game, players must balance their desire for high clustering coefficient against their edge expenditures. The utility of v in the game is defined to be $\textit{utility}(v) = CC(v) - \alpha \cdot \textit{out-deg}(v)$. When the edge cost $\alpha \geq 1$ all strategies for a node v are dominated by the strategy to purchase no edges at all, so we will assume from now on that $0 < \alpha < 1$. Most of our results will consider the natural case in which α is a constant not depending on the population size n — forming edges has a fixed cost — though we will occasionally discuss cases where α diminishes with increasing n. Some of our results will also depend on α being a rational number.

As in much of the related literature, our main interest in this paper is to study the properties of the *pure* Nash equilibrium (NE) networks of the CC game. For some of our results we shall slightly refine this notion to exclude some degenerate cases and thus focus on the interesting ones. Note that the empty network (no edge purchases) is a trivial NE with zero social welfare (total utility) that we will omit from consideration.

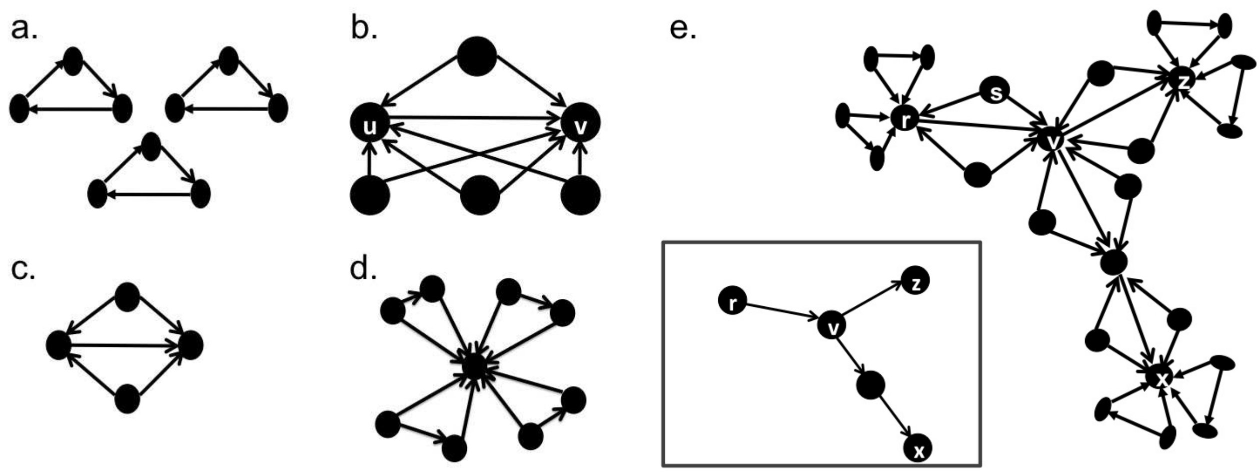

Figure 1: A variety of Nash Equilibrium networks of the CC-Game: Disjoint Triangles NE (a), Popular Victims NE (b and c), Triangular Hub and Spokes NE (d), Binary Tree-Like NE (e).

We also ask that players who purchase edges have non-zero utility. Note that (at least) zero utility can always be obtained by purchasing *no edges*. This condition demands that the action taken by only a subset of the players (those buying edges) be better than only one of their many alternatives (buying no edges), and even then only in the case that the latter gives zero utility. It is thus a considerable weakening of the standard notion of a *strict* Nash equilibrium. We next codify these restrictions:

Definition 1 *A* non-degenerate NE *is a non-empty, pure Nash Equilibrium of the CC game in which out-deg*$(v) \geq 1$ *implies utility*$(v) > 0$ *for all players* v.

The *social welfare* of a given network is defined as the sum of all players' utilities. The (non-degenerate) Price of Anarchy (PoA) is defined as the ratio of the highest social welfare of any directed network with n nodes to the worst social welfare of any non-degenerate NE.

4 A (Partial) Catalog of CC Game Nash Equilibria

We begin by constructing a number of families of non-degenerate NE of the CC game, focusing primarily on the network topologies that can arise at equilibrium. Each of these families is defined for arbitrarily large population size n, and has social welfare scaling linearly with n. We do not propose this catalog to be exhaustive; indeed it is interesting to see the diversity of structures that can arise at equilibrium, and we suspect there are others. Subsequent sections are devoted to the study of general properties of non-degenerate NE.

The first three constructions below are sufficiently simple that their equilibrium proofs can be established by straightforward calculations that we omit. We do provide the equilibrium proof for our last, and richest, construction.

Disjoint Triangles NE. Perhaps the simplest non-degenerate NE consists of $n/3$ disjoint triangles. The nodes in each group form a triangle by purchasing one edge each (Figure 1a). Clearly this structure is a non-degenerate NE for any $0 < \alpha < 1$; for n divisible by 3 it also maximizes the social welfare, a fact we shall use throughout.

Popular Victims NE. This non-degenerate NE shows a case where the most "popular" (highest degree) nodes suffer the lowest utility. Let $n \geq 4$. The construction is as follows: a player u connects to a player v, and each other node connects directly to both u and v by purchasing two edges (Figure 1b). When the edge cost is inversely proportional to n, $\alpha = \frac{2}{n-1} - \epsilon$, for any $\epsilon > 0$, this network is a non-degenerate equilibrium.

To see this, notice that all players other than u and v are playing their best responses and get a positive utility provided $\alpha < \frac{1}{2}$. Node v cannot improve its utility since all nodes are connected to it. Last, if $\alpha < \frac{2}{n-1}$, u wouldn't want to remove the edge it purchased to v and therefore is playing its best response. Furthermore, u is getting a positive utility.

Note that this network is "paid for" by low-degree vertices, all of whom enjoy high utility, while the high-degree victims u and v suffer low utility. We shall show later that in fact this is a property of all non-degenerate NE.

Triangular Hub and Spokes NE. Consider the network shown in Figure 1d; it is easily verified that for edge cost $\alpha = \frac{1}{2} - \epsilon$, for any $\epsilon > 0$ this is a non-degenerate NE. Furthermore, this construction can be scaled up to make the "hub" node have arbitrarily high degree at the same (constant) edge cost, and disjoint copies of this construction of different size can be combined to form new non-degenerate NE. In this fashion we may create non-degenerate NE whose (total) degree distributions are effectively unconstrained.

Binary Tree-Like NE. We next construct a large family of non-degenerate NE obtained by the following construction. We take any rooted, directed binary tree T (with edges always oriented towards the leaves), where the root has out-degree of one, and replace each directed edge in T with a local gadget of the type given in Figure 1c. As an example of the construction, consider starting with the rooted, directed tree T on five vertices shown in Figure 1e (inset). The resulting network $G(T)$ is given in Figure 1e.

It is worth emphasizing that this construction yields a rather rich family of non-degenerate NE with a variety of asymmetries possible, which is somewhat unusual in the network formation game literature. At one extreme it contains connected, small diameter networks (constructed from balanced binary trees) and on the other extreme it contains connected, large diameter networks (constructed from path-like graphs). Since the argument that the construction does indeed yield NE is considerably more involved than for our previous examples, we give a formal theorem and proof below. The basic intuition is that the constructed networks only have constant-sized neighborhoods of dense connectivity (the gadgets). Thus by purchasing additional edges, a node can only linearly increase its triangle count, while the denominator of its clustering coefficient grows quadratically. Furthermore, the gadget edges purchased in the construction can be shown to be beneficial if edge cost is small enough, since such edges have the unique property of contributing two triangles.

Theorem 1 *For any rooted, directed binary tree T where the root has out-degree of one, let $G(T)$ be the directed network obtained by the construction described above. Then for any edge cost that is smaller than some constant independent of network size, $G(T)$ is a non-degenerate pure NE of the CC game.*

Proof: We start by noticing that any new node that is not part of the original tree (namely, a node in a newly capped triangle such as node s) gets a maximum utility when $\alpha < 1/2$: each such node gets a maximum clustering coefficient value of one but only pays for the minimum number of edges (at most two) that makes it's clustering coefficient non zero. We therefore focus only on nodes in the constructed graph that also appear in the original tree (such as nodes r, v, x, z in figure 1e). First we shall consider internal tree nodes; take an arbitrary internal node of T, namely, a node that is not the root or a leaf of the tree T. For example, in the network of Figure 1e, v is such a node. We denote such a node by b. Denote by $1 \leq d \leq 2$ the out degree of b. By the construction b has a total degree of $3d + 3$ (where d is its out-degree in T), an in-degree of $2d + 3$, and an out-degree of d. Moreover, $\Delta(b)$ equals $2d + 2$. Denote the set of d edges purchased by b by E_b. Alternatively, b could have purchased some other set of k edges. What is the best b can do when purchasing k edges?

First, when connecting to gadgets it is always best to connect to consecutive gadgets (namely, gadgets that share a common node). In this way we can 'save an edge' since we already get to be connected to one node in that gadget. Next, it is always at least at good to connect to all nodes in a gadget before connecting to nodes of a consecutive gadget. The reason is that for each new edge we connect to a gadget with, we get an extra of two to b's triangle count. Last, a gadget adjacent to b is always preferable to any other gadget since it gives an addition of two to the triangle count of b. For an illustration see for example the node v and the edge (v, z) in Figure 1e; by purchasing this edge v increases its triangle count by two. Thus the best strategy for b when purchasing $k \leq d$ edges is to keep k out of the d purchased edges in E_b. This strategy is worse than the alternative strategy of keeping all edges in E_b if

$$\frac{2+2k}{\binom{2d+3+k}{2}} - k\alpha \leq \frac{2+2d}{\binom{3d+3}{2}} - d\alpha.$$

For each $k < d$ this translates into an upper bound constraint on α in terms of d:

$$\alpha \leq \frac{1}{d-k}\left(\frac{2+2d}{\binom{3d+3}{2}} - \frac{2+2k}{\binom{2d+3+k}{2}}\right).$$

We next show that the RHS of the above inequality is a strictly positive number which is less than one so we indeed get a realizable constraint on the value of α. First, the RHS is clearly less than one. Next, the RHS is strictly positive iff, after simple algebraic manipulations, $k^2 - k(5d+1) + 4d^2 + d > 0$. Notice that the LHS expression in the last inequality is a parabola in k (thinking of d as fixed); moreover, when $k = 0$ and $k = d-1$ the LHS expression is strictly positive and when $k = d$ the LHS expression equals zero. Thus for any $0 \leq k < d$ the expression is indeed strictly positive.

Next, by a previous discussion, when $k > d$, the best b can do is to keep the edges in E_b and connect to as many consecutive gadgets as possible, getting an addition of five to its triangle count for each three additional purchased edges (and an addition of one more if k mod 3 is one and of two if k mod 3 is two).

We write k as $k = d + 3i + j$, where i, j are natural numbers and $j \leq 2$. We need to show that the utility of this strategy is inferior to the one b uses in G . By the previous discussion b can get to participate in an extra of $5i$ triangles if $j = 0$, and an extra of at most $5i + j + 1$ triangles if $1 \leq j \leq 2$. Therefore if $j = 0$ it suffices to show that

$$\frac{2+2d+5i}{\binom{3d+3+3i}{2}} - (d+3i)\alpha \leq \frac{2+2d}{\binom{3d+3}{2}} - d\alpha,$$

and for $j \geq 1$ it suffices to show that

$$\frac{2+2d+5i+j+1}{\binom{3d+3+3i+j}{2}} - (d+3i+j)\alpha \leq \frac{2+2d}{\binom{3d+3}{2}} - d\alpha.$$

By simple algebra both inequalities hold for any non-negative α (see lemma 1 in Appendix A for a proof).

We are therefore left with the d constraints we had before. To satisfy these constraints it suffices to have α scale like $\frac{1}{d^4}$). Since d is at most two, α can be taken to be any number smaller than an appropriate constant that does not depend on network size.

Next, we analyze the strategies available for a leaf node. We note that a leaf node does not purchase any edges, has an in-degree of three, and is part of two triangles. Purchasing $k = 3i + j$ edges, where i, j are natural numbers and $j \leq 2$, is not profitable for a leaf node; the maximum utility a leaf node can achieve with such k edges is not bigger than what the node gets without purchasing any edges, since by simple algebra (see lemma 2 in Appendix A)

$$\frac{2+5i+j+1}{\binom{3+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}}$$

for all such integers i, j and non-negative α.

Finally, we analyze the strategies for the root. The root bought exactly one edge and has an in-degree of three. Removing this edge yields a lower utility if

$$0 \leq \frac{2}{\binom{3}{2}} - \alpha$$

which holds for $\alpha \leq 2/3$. Next, keeping the edge originally bought and purchasing additional edges is analyzed similarly to previous cases: purchasing $k = 1 + 3i + j$ edges in total (namely, augmenting the current edge with $3i + j$ additional edges), where i, j are natural numbers and $j \leq 2$ is not beneficial; the root's utility would be not be bigger than what the root currently gets, since by simple algebra (see lemma 3 in Appendix A)

$$\frac{2+5i+j+1}{\binom{3+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}} - \alpha$$

for all such integers i, j and non-negative α.

Last, the root could have removed the edge it originally bought and instead could have purchased $k = 3i+j$ other edges where $i \geq 0, j \leq 2$ are natural numbers. Purchasing such k edges is non profitable to the root since by simple algebra (see lemma 4 in Appendix A)

$$\frac{5i+j+1}{\binom{2+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}} - \alpha$$

for all such integers i, j and non-negative α.

Thus by demanding that edge cost α satisfies all the equilibrium inequalities above, which can be done by setting α to a constant not depending on the network size, we get that the set of edge purchases of all nodes is indeed an equilibrium.

∎

5 General Properties of CC Game Nash Equilibria

Given the apparent diversity and potential asymmetry of the NE of the CC game, what general statements might we hope to make about their topological and utility properties? Certain very basic and crude characterizations are easily obtained — for instance, the fact that any NE has at most $\frac{n}{\alpha}$ edges follows from the fact that each node can purchase at most $\frac{1}{\alpha}$ edges at equilibrium since all utilities are non-negative. Notice that this observation does not imply a non-trivial restriction on the total degree or utility of any individual node.

In this section, we prove a considerably stronger characterization motivated by the commonalities in the NE described in the last section. Namely, we prove that any (non-degenerate) NE is paid for by nodes of low total degree and high utility, while high-degree vertices are always victims of low utility. This characterization will then be applied in the following section to obtain non-trivial bounds on the Price of Anarchy for the CC game.

Theorem 2 *Let $0 < \alpha < \frac{1}{2}$. Then in any non-degenerate NE of the CC game:*

- *For any node v, if out-deg$(v) \geq 1$, then deg$(v) < \frac{3}{\alpha}$, and utility$(v) = c(\alpha) > 0$, where the strictly positive constant $c(\alpha)$ depends only on α, and not the population size n. Moreover, when $\frac{1}{\alpha}$ is integral, $c(\alpha) > \frac{\alpha^3}{9}$.* [2] *Thus, vertices purchasing an edge have low total degree and a positive, constant utility.*
- *For any node v with deg$(v) \geq \frac{3}{\alpha}$, utility$(v) < \frac{3}{\alpha(deg(v)-1)}$. Thus, high-degree vertices have low utility.*

Proof: We start by proving the first part of the theorem. Let v be any node in a non-degenerate NE network that purchased an edge and has an in-degree of at least two (the claim is trivially true when in-degree of v is at most one). The upper bound on v's total degree is derived from the fact the v's utility is higher than what it could have gotten by purchasing no edges at all:

$$\frac{\Delta(v)}{\binom{deg(v)}{2}} - \alpha \cdot \textit{out-deg}(v) \geq \frac{\Delta_I(v)}{\binom{\textit{in-deg}(v)}{2}}.$$

Simplifying, we get

$$\Delta_I(v)\left(\frac{2}{deg(v)(deg(v)-1)} - \frac{2}{\textit{in-deg}(v)(\textit{in-deg}(v)-1)}\right) + \frac{\Delta_{I,O}(v)+\Delta_O(v)}{\frac{deg(v)(deg(v)-1)}{2}} \geq \alpha \cdot \textit{out-deg}(v).$$

Since $\frac{2}{deg(v)(deg(v)-1)} - \frac{2}{\textit{in-deg}(v)(\textit{in-deg}(v)-1)} < 0$, we get

$$\frac{\Delta_{I,O}(v)+\Delta_O(v)}{\textit{out-deg}(v)deg(v)(deg(v)-1)} > \frac{\alpha}{2}.$$

[2] A similar bound holds for "simple" rational α; see the proof.

By using $\Delta_O(v) \leq \binom{out\text{-}deg(v)}{2}$ and $\Delta_{I,O}(v) \leq in\text{-}deg(v) \cdot out\text{-}deg(v)$, we get

$$\frac{in\text{-}deg(v)out\text{-}deg(v) + \frac{out\text{-}deg(v)(out\text{-}deg(v)-1)}{2}}{out\text{-}deg(v)deg(v)(deg(v)-1)} > \frac{\alpha}{2},$$

so $\frac{1}{deg(v)} + \frac{1}{2deg(v)} > \frac{\alpha}{2}$, or alternatively, $deg(v) < \frac{3}{\alpha}$.

Next, we prove a lower bound on v's utility that follows from it being strictly positive (non-degeneracy). Recall that $utility(v) = \frac{\Delta(v)}{\binom{deg(v)}{2}} - \alpha out\text{-}deg(v) > 0$. Since $deg(v) < \frac{3}{\alpha}$, the RHS of the utility expression can only equal one out of a finite number possible of possibilities that depend only on α and not on n. In particular, for each α we can choose the worst possible value that still renders $utility(v)$ strictly positive. We denote that value by $c(\alpha)$.

Furthermore if $\frac{1}{\alpha}$ is integral, by taking a common denominator the left hand-side can be written as a strictly positive numerator divided by a denominator of $\frac{1}{\frac{1}{\alpha}\binom{deg(v)}{2}}$. Using $deg(v) < \frac{3}{\alpha}$, we get that v's utility is bigger than $\frac{\alpha^3}{9}$. (More generally note that if $\alpha = p/q$ for integers $p < q$ and thus rational, a similar argument yields a lower bound of $\frac{\alpha^3}{9p}$ on the utility, and thus "simple" α give constructive lower bounds.)

We next prove the second part of the theorem. Consider a node v with a total degree of at least $\frac{3}{\alpha}$. We saw earlier that a node that purchased edges has a degree of less than $\frac{3}{\alpha}$ so v could not have purchased edges at all. Moreover, a node u that purchased an edge to v has degree less than $\frac{3}{\alpha}$ and so v is part of less than $\frac{3}{\alpha}$ joint triangles with u. Therefore the total triangle count of node v is less than $\frac{1}{2}d \cdot \frac{3}{\alpha}$. Thus, v's utility is less than $\frac{\frac{1}{2}d\frac{3}{\alpha}}{\binom{d}{2}} = \frac{3}{\alpha(d-1)}$. ■

6 The Price of Anarchy

As has been mentioned, a disjoint union of triangles is a maximum social welfare NE, whereas all the specific families of NE given in Section 4 have a social welfare growing linearly with the population size n. In this section we prove that the non-degenerate Price of Anarchy is upper bounded by a function depending only on α, and not on n, for all $\alpha < 1/2$, and give an explicit expression for the upper bound when α is a "simple" rational (small numerator). This turns out to be a fairly straightforward consequence of the characterization given in Theorem 2.

Theorem 3 *For edge cost $\alpha \geq \frac{1}{2}$ the non-degenerate Price of Anarchy for the CC game is lower bounded by $\Omega(n(1-\alpha))$, and for edge cost $\alpha < \frac{1}{2}$ it is upper bounded by an expression that depends only on α. Moreover, when $\frac{1}{\alpha}$ is integral the Price of Anarchy is upper bounded by $\frac{36(1-\alpha)}{\alpha^4}$* [3]

Proof: When $\alpha \geq \frac{1}{2}$ a network containing one triangle (where each node purchases exactly one edge) and $n-3$ isolated nodes is a non-degenerate pure NE. The utility of such a network is $3-3\alpha$ while the network with the optimal utility has utility of at least $(n-2)(1-\alpha)$ (by dividing the nodes into disjoint triangles). Thus, the Price of Anarchy is lower bounded by $\Omega(n(1-\alpha))$.

Now consider the case of $\alpha < \frac{1}{2}$. Notice that in equilibrium each node has a positive total degree since it is part of at least one triangle (otherwise it could have bought edges to both endpoints of an existing edge and improve its utility). Since the degree of each node that purchased edges is at most $\frac{3}{\alpha}$ and all nodes in the network need to be touched by an edge, there must be at least $\frac{n\alpha}{3+\alpha} \geq \frac{n\alpha}{4}$ nodes which purchased edges. Next, we know from Theorem 2 that the utility of each node which purchased edges is lower bounded by some constant function $c(\alpha)$ which depends only on α and not n, so the utility of any non-degenerate NE is at least $\frac{n\alpha}{4}c(\alpha)$. Since the welfare of any network is at most $n(1-\alpha)$, the Price of Anarchy is upper bounded by $\frac{4(1-\alpha)}{c(\alpha)\alpha}$. In particular, when $\frac{1}{\alpha}$ is integral we know that $c(\alpha) > \frac{\alpha^3}{9}$ and the Price of Anarchy is upper bounded $\frac{36(1-\alpha)}{\alpha^4}$. As in the proof of Theorem 2, a similar lower bound holds for "simple" (small numerator) rational α. ■

[3] A similar bound holds for "simple" (small numerator) rational α.

While Theorem 3 upper bounds the non-degenerate Price of Anarchy independent of the population size n for $\alpha < 1/2$, it leaves open the question of the exact dependence on α and whether it is even real or not. Indeed, all specific constructions in Section 4 have a constant Price of Anarchy independent of α, even when α is a small numerator rational. We leave the resolution of this dependence as an open problem.

It is natural to ask how robust the results we have described so far are with respect to modifications of the equilibrium notion — especially in light of the results in the following section, where we will prove that even approximate best-response computations for the CC game are intractable. Indeed, it is for similar reasons that in other network formation games, researchers often consider weaker notions of equilibrium, such as *link stability* (which asks only that players cannot improve their utilities by adding or dropping a single edge purchase).

Notice that an equilibrium concept resilient only to the addition or removal of a single one edge already has a Price of Anarchy of $\Omega(n(1-\alpha))$ for any edge cost, since a network with one triangle and many isolated nodes is then in equilibrium no matter how small α is (a single edge purchase can never help). However, define *k-stability* to be the equilibrium concept in which players cannot benefit by switching from their assigned edge purchase set S to any other edge purchase set S' for which the symmetric set difference $|S - S'| \leq k$. (Thus standard link stability corresponds to 1-stability.) For any fixed value of k, computing best responses under k-stability becomes a computationally tractable problem, and for $k \geq 2$, all of our general results can be shown to hold under this notion as well:

Theorem 4 *For all $k \geq 2$, Theorems 2 and 3 remain true when we replace NE by k-stability.*

The proof is omitted, but mainly involves technical modifications of the proof of the first part of Theorem 2 to consider the utility effects of dropping only the most beneficial edge purchases, rather than all edge purchases.

We end by noting that a low PoA implies that the average vertex clustering coefficient is high.

Corollary 1 *For edge cost $\alpha < \frac{1}{2}$ the average vertex clustering coefficient grows at least as some function $g(\alpha)$ independent of the network size, while the network's overall edge density goes to zero at a rate smaller or equal to $\frac{2\alpha}{n-1}$.*

7 Intractability of Best Responses

A natural question that arises in many complex network formation games is how difficult it can be to compute best responses, which would seem a prerequisite to reaching NE dynamically; for instance, best-response computation was shown to be NP-hard to compute for a routing formation game [10]. Here we show that best responses in the CC game are intractable even to approximate, thus motivating the weaker notion of k-stability in the last section.

Theorem 5 *Given a directed graph G and a node v in G (where G represents the edge purchases of the other nodes), the edge cost α (encoded as a rational number), and an integer $f \geq 1$, computing a strategy (set of edge purchases) for v with CC game utility at least $\frac{1}{f}$ of the best-response utility is not polynomial time computable, unless $P = NP$.*

Proof:

The proof is by reduction from the following clique-like problem that we call the $(k-1, k)$-Clique problem: Given an undirected graph H and a number k, decide whether H has a clique of size $s \in \{k-1, k\}$. The $(k-1, k)$-Clique problem can easily be shown to be NP-hard.

We now provide the reduction details. Given a graph H for the $(k, k-1)$-Clique problem we construct the graph G by first replacing each edge $\{a, b\}$ by two directed edges (a, b) and (b, a). We next add to the graph new nodes $u_1, u_2, \ldots, u_t$ for $t = k-2$, as well as one distinguished node v. We next connect each node u_i to v (so each u_i get an out-degree of one). Note that node v has out-degree of zero. The graph G represents the set of all edge purchases of all nodes excluding node v. We will design a value for the edge cost α such that v can achieve a positive utility if and only if the set of nodes it purchased edges to form a clique of size of either $k-1$ or k. Thus, any approximation to best responses that runs in polynomial time can then be used to find such a clique in polynomial-time.

Denote the number of nodes in H by n. For the rest of the proof we will focus on designing the required value for α. This will be done by writing an expression for the utility of v and optimize it appropriately. Without loss of generality we assume that the size of the maximum clique in the original graph H is at least three so $t \geq 1$.

Instead of working directly with v's utility function $\mathit{utility}(v)$ we shall work with an analytically more convenient function which we call the *idealized average utility*. In order to define that function we first compute what is the average utility to v by purchasing exactly $l \geq 2$ edges (buying less than two edges yields a non-positive utility). Clearly all purchased edges are connecting v to nodes that also appear in H (all nodes u_i already connect to v), and v's average utility is $\frac{\mathit{utility}(v)}{l} = \frac{m}{l\binom{l+t}{2}} - \alpha$, where m is the number of edges between the nodes v purchased edges to (namely the number of triangles v is part of). We now define the *idealized average utility* as the maximum value, as a function of l, that the average utility of v can be,

$$\mathit{ideal}_l(v) = \frac{\binom{l}{2}}{l\binom{l+t}{2}} - \alpha.$$

Simplifying the formula first we get $\mathit{ideal}_l(v) = \frac{l-1}{(l+t)(l+t-1)} - \alpha$. In order to extend the function definition over all real numbers l, we shall refer to the function's latter simplification as its definition.

Next we find for which value of l the idealized average utility is maximized by finding the fixed points of the function. Computing the derivative with respect to l and setting it to zero gives $l = \pm(1 + \sqrt{t^2 + t - 1})$. Clearly $l_0 = 1 + \sqrt{t^2 + t - 1}$ is a local maximum of the idealized average utility function (since at $l = 1$ the function equals $-\alpha$ and as l goes to infinity the idealized average utility function converges to $-\alpha$ from above). Therefore the idealized utility is strictly monotonically increasing from $l = 1$ to l_0 and then becomes strictly monotonically decreasing. Thus either $l = t + 1$ or $l = t + 2$ is the integer with the highest idealized utility value among all positive integers.

For both $l = t + 1$ and $l = t + 2$ the idealized utility's value is the same: $\mathit{ideal}_{t+1}(v) = \mathit{ideal}_{t+2}(v) = \frac{1}{2(2t+1)} - \alpha$. Also note that $\mathit{ideal}_t(v) = \frac{t-1}{(2t)(2t-1)} - \alpha$, and $\mathit{ideal}_{t+3}(v) = \frac{t-2}{(2t+3)(2t+2)} - \alpha$.

Next we shall set the edge cost such that the following claim hold: $\mathit{ideal}_l(v) > 0$ if and only if $l \in \{t + 1, t + 2\}$. To do so we define

$$\epsilon = \frac{1}{2} \cdot \min\{(\mathit{ideal}_{t+1}(v) - \mathit{ideal}_t(v)), (\mathit{ideal}_{t+2}(v) - \mathit{ideal}_{t+3}(v)), \frac{1}{2(2t+1)}, \frac{1}{n(n+k)^2}\}$$

and set α to be,

$$\alpha = \frac{1}{2(2t+1)} - \epsilon.$$

Notice that α is well defined since ϵ does not depend on α, α is assigned a value between zero and one, and that α is indeed polynomially encodable in the graph size of H (see lemma 5 in Appendix A for a formal proof).

We now show that the claim stated above holds. Let $l \notin \{t+1, t+2\}$. If $1 \leq l < t+1$ then $\mathit{ideal}_l(v) = \frac{l-1}{(l+t)(l+t-1)} - \alpha = \frac{l-1}{(l+t)(l+t-1)} - \frac{1}{2(2t+1)} + \epsilon$ which can be written as $(\frac{l-1}{(l+t)(l+t-1)} - \alpha) - (\frac{1}{2(2t+1)} - \alpha) + \epsilon \leq \mathit{ideal}_l(v) - \mathit{ideal}_{t+1}(v) + 1/2(\mathit{ideal}_{t+1}(v) - \mathit{ideal}_t(v))$ which is at most, $\mathit{ideal}_t(v) - \mathit{ideal}_{t+1}(v) + 1/2(\mathit{ideal}_{t+1}(v) - \mathit{ideal}_t(v)) \leq 0$. A similar argument shows that $\mathit{ideal}_l(v) \leq 0$ for $l > t + 2$. Next, if $l \in \{t + 1, t + 2\}$ then $\mathit{ideal}_l(v) = \epsilon > 0$.

We are now ready to finish-up the proof. We shall make use of the fact that $\mathit{utility}(v) \leq l \cdot \mathit{ideal}_l(v)$ and equality holds if and only if the l nodes v buys edges to form a clique. By the previous discussion $\mathit{utility}(v) \leq 0$ when v connects to $l \notin \{t+1, t+2\}$ nodes. What about connecting to $l \in \{t+1, t+2\}$ nodes? If those nodes do not form a clique, namely, the number of edges between those nodes is at most $\binom{l}{2} - 1$ then $\mathit{utility}(v) < 0$, since ϵ is too small, namely, $\epsilon \leq \frac{1}{n(n+k)^2} < \frac{1}{l\binom{l+t}{2}}$. However, if $l \in \{t + 1, t + 2\}$ and the l nodes v connects to form a clique, then by connecting to those nodes $\mathit{utility}(v) = \epsilon > 0$.

Thus, any approximation to the optimal utility must return the nodes of a clique of size either $t + 2 = k$ or $t + 1 = k - 1$, given that such clique exists in H .

∎

One way to deal with the inapproximability of best response is to focus on computing best responses under k-stability. Although the problem of computing best response under k stability for each node becomes

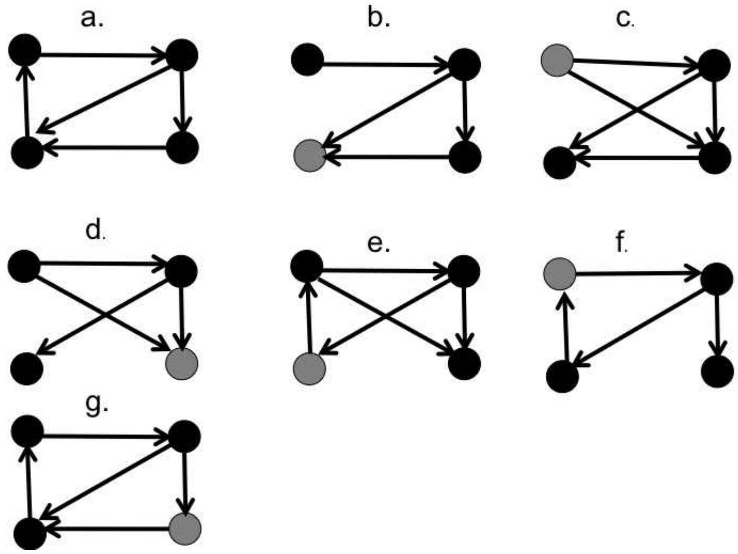

Figure 2: Here we consider the weakest notion of best response, where a node can either add or remove at most one edge to improve its utility. Edge cost can be taken to be any $0 < \alpha < 1$. Figure 3a shows the initial network, and in each consecutive round the network is drawn after the node colored gray played its best-response under 1-stability. The dynamic returns to its initial configuration after six rounds as shown in Figure 3g, so it never converges to a 1-stable network. Furthermore, the same example shows that if, for some $k \geq 1$, at all rounds nodes colored gray played their best-response under k-stability, the dynamic doesn't converge to a k-stable network.

tractable for fixed values of k, the corresponding dynamic doesn't always converge to a k-stable network, as shown in Figure 3. Therefore there is no simple solution to the inapproximability of best-responses.

8 Game Variants

We end by shortly describing our results for two natural variants of the CC game.

An alternative measure of clustering to the commonly used clustering coefficient is the total number of triangles in a network — namely, the *unnormalized* clustering coefficient. Social networks are known to have an exceptionally high triangle count compared to what one would expect in a random Erdös-Renyi network with the same average degree [12]. In the unnormalized version of the CC game, the utility of a node is the number of triangles it is part of minus the edge purchases. We briefly survey the similarities and differences between the new game, which we call the UCC game, and the CC game presented in the paper.

Like the CC game, the UCC game enjoys a variety of equilibrium networks. Notice that an edge cost bigger than one is natural for the UCC game since a node's utility can scale with the network size. Given that, one can show that in contrast to the CC game a clique is a non-degenerate NE of the UCC game for a large range of edge costs. Moreover, any rooted binary tree graph (with a root of out degree one) where we replace all tree edges by cliques of a given size of at least four is a non-degenerate NE for a large range of edge costs. In addition, networks with highly variable degree distribution can be constructed in a similar way to the construction given in the paper for the CC game (by using cliques instead of triangles).

In a non-degenerate equilibrium of the UCC game, each node that purchased an edge has a high degree of at least $\frac{1}{\alpha}$ and utility of at least one. In contrast to the CC-Game, a node that didn't purchase edges can have a high or low degree, and a high utility even when having a high degree.

In contrast to the CC game, the Price of Anarchy of the UCC game depends on the network size and grows as $\Omega(\frac{n}{\alpha})$ for a large range of edge costs. Such a Price of Anarchy ratio is achieved between a clique's welfare to the welfare of a network made of isolated cliques each on α nodes (for an appropriate orientation of clique edges).

As in the CC game, best-response in the UCC is not approximable to any finite factor, unless P = NP.

In contrast to the CC game, for UCC the best-response dynamic converges, and also does so in polynomial time. The reason is that the UCC game is a potential game with a potential function that equals one third of the network's welfare.

We end by mentioning another natural variant of the CC game – a bilateral version of it where a strategy of a player comprises of a set of edge announcements and an edge is created only if both its endpoints announce it. However, under this setting, the set of NE trivializes since any NE network can be shown to be made of a set of isolated triangles and isolated nodes. Moreover, a network maximizing the social welfare is a set of isolated triangles so the PoA grows linearly with the network size.

References

[1] S. Albers, S. Eilts, E. Even-Dar, Y. Mansour, and L. Roditty. On nash equilibria for a network creation game. In *SODA*, pages 89–98, 2006.

[2] N. Alon, E. D. Demaine, M. Hajiaghayi, and T. Leighton. Basic network creation games. In *SPAA*, pages 106–113, 2010.

[3] V. Bala and S. Goyal. A noncooperative model of network formation. *Econometrica*, 68(5):1181–1230, 2000.

[4] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286:509–512, 1999.

[5] B. Bollobás, O. Riordan, J. Spencer, and G. Tusnády. The degree sequence of a scale-free random graph process. *Random Struct. Algorithms*, 18(3):279–290, 2001.

[6] C. Borgs, J. T. Chayes, J. Ding, and B. Lucier. The hitchhiker's guide to affiliation networks: A game-theoretic approach. In *ICS*, 2011.

[7] D. Easley and J. Kleinberg. *Networks Crowds and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010.

[8] E. Even-Dar, M. Kearns, and S. Suri. A network formation game for bipartite exchange economies. In *SODA*, pages 697–706, 2007.

[9] Eyal Even-Dar and Michael Kearns. A small world threshold for economic network formation. In *NIPS*, pages 385–392, 2006.

[10] A. Fabrikant, A. Luthra, E. N. Maneva, C. H. Papadimitriou, and S. Shenker. On a network creation game. In *PODC*, pages 347–351, 2003.

[11] F. Heider. *The Psychology of Interpersonal Relations*. John Wiley & Sons, 1958.

[12] M. O. Jackson. *Social and Economic Networks*. Princeton University Press, 2008.

[13] M. O. Jackson and A. Wolinsky. A strategic model of social and economic networks. In *J. of Economic Theory*, pages 71:44–74, 1996.

[14] C. Johnson and R.P. Gilles. Spatial social networks. In *Review of Economic Design*, pages 5:273–299, 2000.

[15] Jon M. Kleinberg. The small-world phenomenon: an algorithm perspective. In *STOC*, pages 163–170, 2000.

[16] E. Koutsoupias and C. H. Papadimitriou. Worst-case equilibria. In *STACS*, pages 404–413, 1999.

[17] S. Lattanzi and D. Sivakumar. Affiliation networks. In *STOC*, pages 427–434, 2009.

[18] M. Newman, A. L. Barabasi, and D. J. Watts. *The Structure and Dynamics of Networks*. Princeton University Press, 2006.

[19] D. J. Watts. *Small worlds*. Princeton University Press, 1999.

[20] D. J. Watts and S. H. Strogatz. Collective dynamics of 'small-world' networks. *Nature*, 393(6684):440–442, 1998.

A Omitted Proofs

Lemma 1 *Let i, j be natural numbers, $i \geq 0$, $1 \leq j \leq 2$. Also let $1 \leq d \leq 2$, and $\alpha \geq 0$. Then,*

- $$\frac{2+2d+5i}{\binom{3d+3+3i}{2}} - (d+3i)\alpha \leq \frac{2+2d}{\binom{3d+3}{2}} - d\alpha$$

- $$\frac{2+2d+5i+j+1}{\binom{3d+3+3i+j}{2}} - (d+3i+j)\alpha \leq \frac{2+2d}{\binom{3d+3}{2}} - d\alpha$$

Proof: We begin by proving the first inequality. It suffices to show that

$$\frac{2+2d+5i}{\binom{3d+3+3i}{2}} \leq \frac{2+2d}{\binom{3d+3}{2}}.$$

By simple algebra this is equivalent to showing

$$3(3d+2)(2+2d+5i) \leq 2(3d+3+3i)(3d+2+3i)$$

which is equivalent to

$$18d^2 + 30d + 45di + 30i + 12 \leq$$
$$18d^2 + 30d + 36di + 30i + 18i^2 + 12$$

which is equivalent to

$$9di \leq 18i^2,$$

which holds for $d \leq 2, i \geq 0$.

Similarly, for proving the second equality it suffices to prove that

$$9di + 3(3d+2)(j+1) \leq$$
$$18i^2 + 2j(3d+2+3i+j) + 2j(3d+3+3i+j)$$

and it suffices to show that

$$9di + (9d+6)j + (9d+6) \leq 18i^2 + j(12d+10+12i+4j)$$

which holds if

$$9d+6 \leq j(3d+4+12i+4j)$$

or if

$$9d+6 \leq 3d+4+12j+4j$$

which holds for $d \leq 2$ and any $j \geq 1$.

∎

Lemma 2 *Let $k = 3i + j$, where i, j are natural numbers not both zero and $j \leq 2$. Let $1 \leq d \leq 2$, and let $\alpha \geq 0$. Then,*

$$\frac{2+5i+j+1}{\binom{3+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}}$$

Proof: To prove the inequality it suffices to show

$$\frac{2+5i+j+1}{\binom{3+3i+j}{2}} \leq \frac{2}{\binom{3}{2}}$$

for all integers i, j and non-negative α. This is equivalent to showing that

$$3(2+5i+j+1) \leq (3+3i+j)(2+3i+j)$$

which is equivalent to showing that

$$6+15i+3j+3 \leq 6+15i+5j+9i^2+6ij+j^2$$

which is equivalent to

$$3 \leq 2j+9i^2+6ij+j^2$$

which holds when $j \geq 1, i \geq 0$ and also when $i \geq 1, j \geq 0$.

■

Lemma 3 *Let $k = 3i+j$, where i, j are natural numbers not both zero and $j \leq 2$. Let $\alpha \geq 0$. Then,*

$$\frac{2+5i+j+1}{\binom{3+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}} - \alpha.$$

Proof: the proof of lemma 2 covers also this case.

■

Lemma 4 *Let $k = 3i+j$, where i, j are natural numbers not both zero and $j \leq 2$. Let $\alpha \geq 0$. Then,*

$$\frac{5i+j+1}{\binom{2+3i+j}{2}} - k\alpha \leq \frac{2}{\binom{3}{2}} - \alpha$$

Proof: It suffices to show that

$$\frac{5i+j+1}{\binom{2+3i+j}{2}} \leq \frac{2}{\binom{3}{2}} - \alpha.$$

This is equivalent to showing that

$$15i+3j+3 \leq (2+3i+j)(1+3i+j)$$

which is equivalent to

$$15i+3j+3 \leq 9i^2+9i+6ij+3j+j^2+2$$

which is also equivalent to

$$6i+1 \leq 9i^2+6ij+j^2$$

which holds when $j \geq 1, i \geq 0$ and also when $i \geq 1, j \geq 0$. ■

Lemma 5 *Define $\epsilon = \frac{1}{2} \cdot \min\{(ideal_{t+1}(v) - ideal_t(v)), (ideal_{t+2}(v) - ideal_{t+3}(v)), \frac{1}{2(2t+1)}, \frac{1}{(n+k)^2}\}$, and set α to be, $\alpha = \frac{1}{2(2t+1)} - \epsilon$. Then,*

- *α is well defined.*
- *α is assigned a value between zero and one.*
- *α is polynomially encodable in the size of the graph H .*

Proof: We first show that α is well defined. This follows from the fact the ϵ does not depend on α but only on t. $\epsilon = \frac{1}{2} \cdot \min\{(ideal_{t+1}(v) - ideal_t(v)), (ideal_{t+2}(v) - ideal_{t+3}(v)), \frac{1}{2(2t+1)}, \frac{1}{n(n+k)^2}\} = \frac{1}{2} \cdot \min\{(\frac{1}{2(2t+1)} - \frac{t-1}{(2t)(2t-1)}), (\frac{1}{2(2t+1)} - \frac{t-2}{(2t+3)(2t+2)}), \frac{1}{2(2t+1)}, \frac{1}{n(n+k)^2}\}$.

Next we show that $0 \leq \alpha \leq 1$. First, $\alpha = \frac{1}{2(2t+1)} - \epsilon \leq \frac{1}{2(2t+1)} < 1$. Next, $\alpha \geq 0$ iff $\epsilon \leq \frac{1}{2(2t+1)}$ which holds by definition of ϵ. Last notice that ϵ is a defined as a difference between two simple rational functions in variables t, n, k and that $1 \leq t \leq n$, $3 \leq k \leq n$, so ϵ is polynomially encodable, as a numerator followed by a denominator, in the length of n (and so in the size of H).

■