

# Selfish Creation of Social Networks

Davide Bilò<sup>1</sup>, Tobias Friedrich<sup>2</sup>, Pascal Lenzner<sup>2</sup>, Stefanie Lowski<sup>3</sup>, Anna Melnichenko<sup>2</sup>

<sup>1</sup> Department of Humanities and Social Sciences, University of Sassari, Italy

<sup>2</sup> Hasso Plattner Institute, University of Potsdam, Germany

<sup>3</sup> Department of Computer Science, Humboldt-University Berlin, Germany

davide.bilo@uniss.it, tobias.friedrich@hpi.de, pascal.lenzner@hpi.de, lowski@math.hu-berlin.de, anna.melnichenko@hpi.de

## Abstract

Understanding real-world networks has been a core research endeavor throughout the last two decades. Network Creation Games are a promising approach for this from a game-theoretic perspective. In these games, selfish agents corresponding to nodes in a network strategically decide which links to form to optimize their centrality. Many versions have been introduced and analyzed, but none of them fits to modeling the evolution of social networks. In real-world social networks, connections are often established by recommendations from common acquaintances or by a chain of such recommendations. Thus establishing and maintaining a contact with a friend of a friend is easier than connecting to complete strangers. This explains the high clustering, i.e., the abundance of triangles, in real-world social networks.

We propose and analyze a network creation model inspired by real-world social networks. Edges are formed in our model via bilateral consent of both endpoints and the cost for establishing and maintaining an edge is proportional to the distance of the endpoints before establishing the connection. We provide results for generic cost functions, which essentially only must be convex functions in the distance of the endpoints without the respective edge. For this broad class of cost functions, we provide many structural properties of equilibrium networks and prove (almost) tight bounds on the diameter, the Price of Anarchy and the Price of Stability. Moreover, as a proof-of-concept we show via experiments that the created equilibrium networks of our model indeed closely mimics real-world social networks. We observe degree distributions that seem to follow a power-law, high clustering, and low diameters. This can be seen as a promising first step towards game-theoretic network creation models that predict networks featuring all core real-world properties.

## Introduction

Complex networks from the Internet to various (online) social networks have a huge impact on our lives and it is thus an important research challenge to understand these networks and the forces that shape them. The emergence of the Internet has kindled the interdisciplinary field of Network Science (Barabási 2016), which is devoted to analyzing and understanding real-world networks.

Extensive research, e.g. (Albert, Jeong, and Barabási 1999; Barabási and Albert 1999; Broder et al. 2000; Kleinberg 2000; Leskovec, Kleinberg, and Faloutsos 2005; Newman, Barabási, and Watts 2011; Barabási 2016), on real world networks from many different domains like communication networks, social networks, metabolic networks, etc. has revealed the astonishing fact that most of these real-world networks share the following basic properties:

- *Small-world property*: The diameter and average distances are at most logarithmic in number of nodes.
- *Clustering*: Two nodes with a common neighbor have a high probability of being neighbors, i.e., there is an abundance of triangles and small cliques.
- *Power-law degree distribution*: The probability that a node has degree  $k$  is proportional to  $k^{-\beta}$ , for  $2 \leq \beta \leq 3$ . That is, the degree distribution follows a power-law. Such networks are also called *scale-free networks*.

The phenomenon that real world networks from different domains are very similar begs a scientific explanation, i.e., formal models that generate networks with the above properties from very simple rules.

Many such models have been proposed, most prominently the preferential attachment model (Barabási and Albert 1999), Chung-Lu random graphs (Chung and Lu 2002), hyperbolic random graphs (Krioukov et al. 2010; Friedrich and Krohmer 2015) and geometric inhomogeneous random graphs (Bringmann, Keusch, and Lengler 2019). However, all these models describe a purely random process which eventually outputs a network having realistic properties. In contrast, many real-world networks evolved over time by an interaction of rational agents. For example, in (online) social networks (Jackson 2010) the selfish agents correspond to people or firms that choose carefully with whom to maintain a connection. Thus, a model with higher explanatory power should consider rational selfish agents which use and modify the network to their advantage (Papadimitriou 2001).

In game-theoretic models for network formation, selfish agents correspond to nodes in a network. Each agent strategically selects other agents to form a link. The union of all chosen links then determines the edge-set of the created network (Jackson and Wolinsky 1996). The individual goal of each agent is modeled via a cost function, which typically consists of costs for creating links and of a service cost

term, which measures the perceived quality of the created network for the individual agent. For example, the service cost could be proportional to the node's centrality (Fabrikant et al. 2003) or simply to the number of reachable nodes (Bala and Goyal 2000). Any network can be considered as an outcome of such a game and among all possible networks the so-called equilibrium networks, where no agent wants to add or remove links, are particularly interesting since analyzing their structure yields insights into why real-world networks exhibit the above-mentioned properties. Moreover, a game-theoretic model allows measuring the impact of the selfish agent behavior on the whole society of agents.

So far, game-theoretic approaches can explain the small-world property, that is, it has been proven that the diameter of all equilibrium networks is small (Demaine et al. 2012). However, to the best of our knowledge, no known game-theoretic model can also explain the emergence of clustering and a power-law degree distribution.

**Our Contribution** In this paper, we propose and analyze a simple and very general game-theoretic model which is motivated by real-world social networks. Its main actors are selfish agents that bilaterally form costly links to increase their centrality. Hereby, the cost of each link is an arbitrary convex function in the distance of the involved nodes without this link. This naturally models the convention that connecting with a friend of a friend is much easier than to establish and maintain a link with an agent having no common acquaintances. To establish a link, both endpoints have to agree and pay the edge's cost.

We characterize the social optimum and prove the existence of equilibrium networks for our generic model. For this, we provide many structural properties. Moreover, we give (nearly) tight bounds on the diameter, the Price of Anarchy and the Price of Stability that essentially only depend on the cost of closing a triangle and on the cost of maintaining a bridge-edge in the network. This implies that all these values are very low for many natural edge-cost functions. Moreover, as a proof of concept, we show via simulation experiments of our model that a given sparse initial network evolves over time into an equilibrium network having a power-law degree distribution, high clustering and a low diameter. Hence, our model promises to be the first game-theoretic network formation model which predicts networks that exhibit all core properties of real-world networks.

Due to the space constraints some details are omitted. See (Bilò et al. 2020) for the full version of the paper.

**Model and Notation** We consider a model which is related to the bilateral network creation game (Corbo and Parkes 2005). In our model, called *social network creation game* (SNCG), the set of  $n$  selfish agents  $V$  corresponds to the nodes of a network and the agents' strategies determine the edge-set of the formed network  $G$ . More precisely, let  $\mathbf{s} = (S_1, \dots, S_n)$  denote the strategy profile, where  $S_u \subseteq V \setminus \{u\}$  corresponds to agent  $u$ 's strategy, then the jointly created network  $G(\mathbf{s})$  is defined as  $G(\mathbf{s}) = (V, E)$ , with  $E = \{\{u, v\} \mid u, v \in V, u \in S_v, v \in S_u\}$ . And, in-

versely, for any given undirected network  $G = (V, E)$  there exists a minimal strategy vector  $\mathbf{s} = (S_1, \dots, S_{|V|})$  with  $u \in S_v$  and  $v \in S_u$  if and only if  $\{u, v\} \in E$ , that realizes this network, i.e., with  $G = G(\mathbf{s})$ . Hence, we will omit the reference to  $\mathbf{s}$ . Also, we will use the shorthand  $uv$  for the undirected edge  $\{u, v\}$ .

Each agent  $u$  tries to optimize a cost function  $\text{cost}(G, u)$ , which solely depends on the structure of the network  $G$ . In real-world social networks new connections are formed by a bilateral agreement of both endpoints while an existing connection can be unilaterally removed by any one of the involved endpoints. Following this idea, we consider only single edge additions with consent of both endpoints or single edge deletions as possible (joint) strategy changes of the agents. As equilibrium concept we adopt the well-known solution concept called *pairwise stability* (Jackson and Wolinsky 1996). Intuitively, a network  $G$  is pairwise stable if every edge of  $G$  is beneficial for both endpoints of the edge and for every non-edge of  $G$ , at least one endpoint of that edge would increase her cost by creating the edge. More formally,  $G = (V, E)$  is pairwise stable if and only if the following conditions hold:

1. for every edge  $uv \in E$ , we have  $\text{cost}(G - uv, u) \geq \text{cost}(G, u)$  and  $\text{cost}(G - uv, v) \geq \text{cost}(G, v)$ ;
2. for every non-edge  $uv \notin E$ , we have  $\text{cost}(G + uv, u) \geq \text{cost}(G, u)$  or  $\text{cost}(G + uv, v) \geq \text{cost}(G, v)$ ;

where  $G - uv$  (resp.,  $G + uv$ ) denotes the network  $G$  in which the edge  $uv$  has been deleted (resp., added). Created edges are bidirectional and can be used by all agents, but the cost of each edge is equally shared by its two endpoints.

We denote by  $d_G(u, v)$  the distance between  $u$  and  $v$  in  $G = (V, E)$ , i.e., the number of edges in a shortest path between  $u$  and  $v$  in  $G$ . We assume that  $d_G(u, v) = +\infty$  if no path between  $u$  and  $v$  exists in  $G$ . The main novel feature of our model is the definition of the cost of any edge  $uv \in E$ , which is proportional to the distance of both endpoints without the respective edge, i.e., proportional to  $d_{G-uv}(u, v)$ . This is motivated by the fact that, in social networks, the probability of establishing a new connection between two parties is inversely proportional to their degree of separation. More precisely, let  $\sigma : \mathbb{R} \rightarrow \mathbb{R}$  be a monotonically increasing convex function such that  $\sigma(0) = 0$ .<sup>1</sup> The cost of the edge  $uv$  in network  $G$  is equal to

$$c_G(uv) = \begin{cases} \sigma(d_{G-uv}(u, v)) & \text{if } d_{G-uv}(u, v) \neq +\infty, \\ \sigma(n) & \text{otherwise.} \end{cases}$$

We call an edge  $uv$  a  $k$ -edge if  $d_{G-uv}(u, v) = k$ , and a *bridge* if  $d_{G-uv}(u, v) = +\infty$ . If the network is clear from the context, we will sometimes omit the reference to  $G$  and we still simply write  $c(uv)$  to denote the cost of edge  $uv$ . Note that by definition, any bridge in  $G$ , i.e., any edge whose removal would increase the number of connected components of  $G$ , has cost  $\sigma(n) > \sigma(n-1)$  and thus any bridge has higher cost than any other non-bridge edge. The latter

<sup>1</sup> All the results of this paper hold if we replace this constraint by the milder constraint  $\sigma(3) \geq \frac{3}{2}\sigma(2)$ .

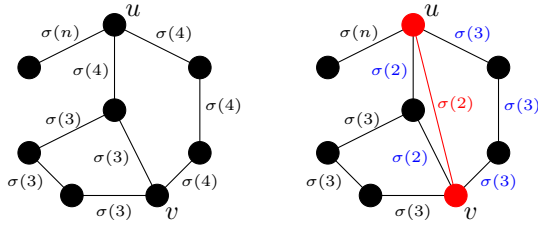


Figure 1: Edge costs before and after the edge  $uv$  is added.

property can be understood as an incentive towards more robust networks. Note, that the addition or removal of an edge in network  $G$  can also influence the cost of other edges in  $G$ . See Figure 1 for an example.

The resulting cost for an agent  $u$  in the network  $G$  is the sum of the cost of all edges incident to  $u$  and the sum of distances to all other agents:

$$\text{cost}(G, u) := \frac{1}{2} \sum_{v \in N_G(u)} c_G(uv) + \sum_{v \in V} d_G(u, v),$$

where  $N_G(u)$  is the set of all neighbors of  $u$  in  $G$ . The quality of the created network is measured by its *social cost*, which is denoted by  $\text{cost}(G) := \sum_{u \in V} \text{cost}(G, u)$  and measures the total cost of all agents.

As in the original bilateral network creation game (Corbo and Parkes 2005), we restrict our study to connected networks, as any pairwise stable non-connected network has an unbounded social cost<sup>2</sup>. Let  $\text{worst}_n$  (resp.,  $\text{best}_n$ ) be the highest (resp., lowest) possible social cost of a pairwise stable (connected) network created by  $n$  agents, assuming that a pairwise stable network with  $n$  agents always exists. Moreover, let  $\text{opt}_n$  be the social cost of a *social optimum*, i.e., a minimum social cost network of  $n$  nodes. Then the *Price of Anarchy (PoA)* (Koutsoupias and Papadimitriou 1999) is defined as  $\max_{n \in \mathbb{N}} \frac{\text{worst}_n}{\text{opt}_n}$  and measures the deterioration of the network's social cost due to the agents' selfishness. The *Price of Stability (PoS)* (Anshelevich et al. 2008) is the ratio  $\max_{n \in \mathbb{N}} \frac{\text{best}_n}{\text{opt}_n}$  and describes the minimal cost discrepancy between an equilibrium and an optimal outcome.

**Related Work** Strategic network formation is a rich and diverse research area and it is impossible to discuss all previous work in detail. Instead, we focus on the models which are closest to our approach.

The SNCG is a variant of the bilateral network creation game (BNCG) (Corbo and Parkes 2005). The BNCG is based on the unilateral network creation game (NCG) (Fabrikant et al. 2003) where edges can be created and must be paid by only one of its endpoints, and the pure Nash equilibrium is used as solution concept. In both models edges have a uniform cost of  $\alpha > 0$ . In a series of papers it was established that the PoS for the NCG is at most  $\frac{4}{3}$  and the PoA is constant for almost all values of  $\alpha$  (Fabrikant et al. 2003; Albers et al. 2006; Demaine et al. 2012; Mihalák and Schlegel

2010; Mamageishvili, Mihalák, and Müller 2013; Bilò and Lenzner 2020; Álvarez and Mesegué 2019). Moreover, it was shown that the diameter of any equilibrium network in the NCG is at most  $2^{\mathcal{O}(\sqrt{\log n})}$  (Demaine et al. 2012) and for many ranges of  $\alpha$  it is constant. In contrast, for the BNCG it was shown the PoA of the BNCG is in  $\Theta(\min\{\sqrt{\alpha}, n/\sqrt{\alpha}\})$  and that a equilibrium networks with a diameter in  $\Theta(\sqrt{\alpha})$  exist (Corbo and Parkes 2005; Demaine et al. 2012). The original NCG was dedicated to model real-world networks like peer-to-peer networks and social networks. However, the main downside of these classical models is that they do not predict a realistic degree distribution or high clustering. A NCG variant was proposed where agents try to maximize their local clustering instead of their centrality (Brautbar and Kearns 2011). This model yields various sparse equilibrium networks with high clustering but these can have a large diameter and a homogeneous degree distribution.

Closer to the SNCG are variants of the NCG with non-uniform edge cost. Models were proposed where the edge cost is proportional to its quality (Cord-Landwehr, Mäcker, and auf der Heide 2014). Edges between certain types of agents have different costs (Meirom, Mannor, and Orda 2014, 2015), and the edge cost depends on the node degree (Chauhan et al. 2017), or the edge costs are defined by an underlying geometry (Bilò et al. 2019). Especially the latter is related to our model, as our model can also be understood as having a dynamically changing underlying geometry which depends on the structure of the current network. Finally, the island connection model (Jackson and Rogers 2005) assumes that groups of agents are based on islands and that the edge cost within an island is much lower than across islands. This yields equilibria with low diameter and high clustering but no realistic degree distribution.

The SNCG incorporates a robustness aspect since bridge-edges are expensive. This fits to a recent trend in the AI community for studying robust network formation (Meirom, Mannor, and Orda 2015; Chauhan et al. 2016; Goyal et al. 2016; Chen et al. 2019; Echzell et al. 2020).

Despite the variety of studied network formation models, to the best of our knowledge, no simple game-theoretic model exists, which predicts a low diameter, a power-law degree distribution and high clustering in its equilibrium networks. We are also not aware of any simulation results in this direction. However, there are two promising but very specialized candidates in that direction. The first candidate, which is particularly tailored to the web graph (Kouroupas et al. 2015), yields directed equilibrium networks that share many features of real-world content networks. The second candidate uses a game-theoretic framework and hyperbolic geometry to generate networks with real-world features. In the network navigation game (Gulyás et al. 2015), agents correspond to randomly sampled points in the hyperbolic plane and they strategically create edges to ensure greedy routing in the created network. It is shown that the equilibrium networks indeed have a power-law degree distribution and high clustering. However, the main reason for this is not the strategic behavior of the agents but the fact that the agents correspond to uniformly sampled points in the hyper-

<sup>2</sup>E.g., any network with no edges and  $n \geq 3$  is pairwise stable.

bolic plane. It is known that the closely related hyperbolic random graphs (Krioukov et al. 2010) indeed have all core properties of real-world networks.

## Properties of Equilibrium Networks

In this section we prove structural properties satisfied by all connected pairwise stable networks that will be useful in proving our main results. We first define some basic notation and provide a nice property satisfied by the function  $\sigma$ . An edge  $e$  of a network  $G$  is a *bridge* if  $G - e$  has at least one more connected component than  $G$ . A connected network that has no bridge is said to be *2-edge-connected*. A *2-edge-connected component* of a network  $G$  is a maximal (w.r.t. node addition) induced subgraph of  $G$  that is 2-edge-connected.<sup>3</sup> The *diameter*  $D$  of a network  $G$  is equal to the length of the longest shortest path in  $G$ , i.e.,  $D = \max_{u,v \in V} d_G(u, v)$ . Finally, we say that an edge  $uv$  of  $G$  is an  *$i$ -edge* if  $d_{G-uv}(u, v) = i$ , where we use the convention  $n$ -edge for a bridge edge.

**Proposition 1.** Fix a positive real value  $x$ . Let  $x_1, \dots, x_k$ , with  $0 \leq x_i \leq x$ , be  $k \geq 2$  positive real values and let  $\lambda_1, \dots, \lambda_k$ , with  $\lambda \in [0, 1]$ , such that  $x = \sum_{i=1}^k (\lambda_i x_i)$ . Then  $\sigma(x) \geq \sum_{i=1}^k (\lambda_i \sigma(x_i))$ .

In the next statement we claim that nodes can be incident to at most one expensive edge. Hence, the number of such edges is limited.

**Proposition 2.** In any pairwise stable network, any node has at most one incident edge of cost at least  $\sigma(4)$ . If  $2\sigma(2) \leq \sigma(3)$  holds, any node in a pairwise stable network has at most one incident edge of cost at least  $\sigma(3)$ .

Next, we establish that all pairwise stable networks contain at most three bridges.

**Proposition 3.** Any pairwise stable network contains at most three bridges.

The following proposition shows an upper bound of the diameter of any pairwise stable network that only depends on the cost of edges which close a triangle.

**Proposition 4.** The diameter of any pairwise stable network is at most  $\sigma(2) + 2$ .

*Proof.* Consider a pairwise stable network  $G$  of diameter  $D$ . Let  $v_0, v_1, \dots, v_D$  be a diametral path of  $G$ . Consider the addition of the edge between  $v_{\lfloor D/2 \rfloor - 1}$  and  $v_{\lfloor D/2 \rfloor + 1}$  to network  $G$ . Each node  $v_0, \dots, v_{\lfloor D/2 \rfloor - 1}$  becomes 1 unit closer to  $v_{\lfloor D/2 \rfloor + 1}$ ; similarly, each node  $v_{\lfloor D/2 \rfloor + 1}, \dots, v_D$  becomes 1 unit closer to  $v_{\lfloor D/2 \rfloor - 1}$ . In both cases, the distance cost of the considered agent decreases by at least  $\lfloor D/2 \rfloor$ . Since the network is pairwise stable, both agents  $v_{\lfloor D/2 \rfloor - 1}$  and  $v_{\lfloor D/2 \rfloor + 1}$  have no incentive in buying the considered edge. Therefore,  $\sigma(2)/2 - \lfloor D/2 \rfloor \geq 0$  from which we derive  $D \leq \sigma(2) + 2$ .  $\square$

<sup>3</sup>The subgraph of  $G$  induced by a node set  $U \subseteq V$  is a subgraph whose node set is  $U$  that, for any two nodes  $u, v \in U$ , contains the edge  $uv$  if  $uv$  is also an edge of  $G$ .

Finally, we prove an upper bound on the cost of non-bridge edges. This implies that all pairwise stable networks contain only small minimal cycles, i.e., cycles where the shortest path between two nodes in the cycle is along the cycle.

**Proposition 5.** In a pairwise stable network, for all  $k \notin \{2, 3, n\}$ , the cost of any  $k$ -edge is  $\sigma(k) < n\sigma(2)$ . If  $\sigma(2) \leq \frac{1}{2}\sigma(3)$  holds, for all  $k \notin \{2, n\}$ , the cost of any  $k$ -edge is  $\sigma(k) \leq n\sigma(2)$ .

## Equilibrium Existence and Social Optima

The *clique graph* of  $n$  nodes is denoted by  $K_n$ . A *fan graph*  $F_n$  with  $n$  nodes consists of a star with  $n - 1$  leaves  $v_0, \dots, v_{n-2}$  augmented with all the edges of the form  $v_{2i}v_{2i+1}$ , for  $i = 0, \dots, \lfloor \frac{n-2}{2} \rfloor$ , where all indices are computed modulo  $n - 2$  (see Figure 2 for examples). In other words,  $F_n$ , with  $n$  odd, is a star augmented with a perfect matching w.r.t. the star leaves,<sup>4</sup> while  $F_n$ , with  $n$  even, consists of  $F_{n-1}$  augmented with an additional node that is connected to the star center and any star leaf. Clique graphs and

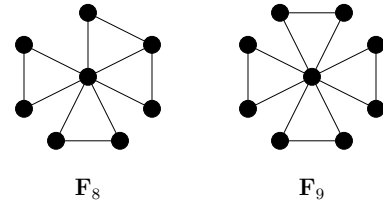


Figure 2: Two examples of fan graphs.

fan graphs play an important role since, as we will prove, the former are social optima when  $\sigma(2) \leq 2$ , while the latter are social optima when  $\sigma(2) \geq 2$ . Furthermore, we also show that clique graphs are pairwise stable whenever  $\sigma(2) \leq 2$ , while (almost) fan graphs are pairwise stable whenever  $\sigma(2) \geq 2$ .

**Theorem 1.** If  $\sigma(2) < 2$ , then  $K_n$  is the unique social optimum. If  $\sigma(2) > 2$ , then  $F_n$  is the unique social optimum. Finally, if  $\sigma(2) = 2$  any network of diameter 2 and containing only 2-edges is a social optimum.<sup>5</sup>

Now we prove the existence of pairwise stable networks. For this we consider a modified fan graph  $F'_n$  that is equal to  $F_n$  if  $n$  is odd. If  $n$  is even,  $F'_n$  consists of  $F_{n-1}$  and one additional node connected to the center.

**Theorem 2.** For  $\sigma(2) \geq 2$ , a the modified fan graph  $F'_n$  is a pairwise stable network, otherwise a clique is the unique pairwise stable network.

## PoA and PoS

Here we prove upper and lower bounds to the PoA and PoS.

**Theorem 3.** The PoA of the SNCG is in  $O\left(\min\{\sigma(2), n\} + \frac{\sigma(n)}{n \max\{\sigma(2), n\}}\right)$ . For the class of 2-edge-connected networks the PoA is in  $O(\min\{\sigma(2), n\})$ .

<sup>4</sup>In the literature, this graph is also known as *friendship graph* or *Dutch windmill graph*.

<sup>5</sup>Hence, for  $\sigma(2) = 2$ ,  $K_n$  and  $F_n$  are also social optima.

*Proofsketch.* By Theorem 1 and Theorem 2, we only need to focus on the case  $\sigma(2) \geq 2$ . Indeed, when  $\sigma(2) < 2$ ,  $\mathbf{K}_n$  is the unique pairwise stable network as well as the unique social optimum and, therefore, the PoA is equal to 1.

For the rest of the proof we assume that  $\sigma(2) \geq 2$ . By Theorem 1,  $\mathbf{F}_n$  is a social optimum of cost  $\Omega(n^2 + \sigma(2)n) = \Omega(n \max\{\sigma(2), n\})$ . Consider a pairwise stable network  $G$  of maximum social cost for a given number of nodes  $n$ . Let  $D$  be the diameter of  $G$ . A trivial upper bound for the distance cost of the network is  $n(n-1) \cdot D$ . By Proposition 4, the network diameter is at most  $\sigma(2) + 2$ , hence the distance cost of  $G$  is at most  $(\sigma(2) + 2) \cdot n(n-1)$ .

Now we will show an upper bound for the edge cost. Let  $k_i$  denote the number of  $i$ -edges in  $G$ . By Proposition 3,  $G$  has at most 3 bridges. Hence, for any pairwise stable network we have that  $k_n \leq 3$ ; if the network is additionally 2-edge-connected, then  $k_n = 0$ . We consider two cases, depending on whether  $2\sigma(2) \leq \sigma(3)$  or not.

We consider the case  $2\sigma(2) \leq \sigma(3)$ . By Proposition 2, each node has at most one incident  $i$ -edge where  $3 \leq i < n$ . Moreover, by Proposition 5,  $\sigma(i) \leq n\sigma(2)$  for any  $i \geq 3$ . Then the overall edge-cost of the network is at most

$$\begin{aligned} k_2 \cdot \sigma(2) + \sum_{i=3}^{n-1} (\sigma(i) \cdot k_i) + k_n \sigma(n) \\ \leq \left( \frac{n(n-1)}{2} - \sum_{i=3}^{n-1} k_i \right) \cdot \sigma(2) + n\sigma(2) \sum_{i=3}^{n-1} k_i + k_n \sigma(n) \\ \leq \sigma(2) \cdot \frac{n(n-1)}{2} + (n-1)\sigma(2) \cdot \frac{n}{2} + k_n \sigma(n) \\ \leq \sigma(2)n^2 + k_n \sigma(n). \end{aligned}$$

Therefore,  $\text{PoA} \in O\left(\min\{\sigma(2), n\} + \frac{\sigma(n)}{n \max\{\sigma(2), n\}}\right)$ , while for the class of 2-edge-connected networks, the PoA is in  $O(\min\{\sigma(2), n\})$ .

If  $2\sigma(2) > \sigma(3)$ , the same upper bound for the edge cost can be proven in a similar way. The statement follows.  $\square$

It is worth noticing that the high inefficiency of worst case pairwise stable networks in Theorem 3 follows from the existence of bridges in a network. The PoA is much better in bridge-free pairwise stable networks. Such networks can for example evolve via edge additions starting from a 2-edge-connected network. A real-world example for this would be co-authorship networks of authors with at least two papers.

We now prove lower bounds on the PoA. We start with the construction of a pairwise stable 2-edge-connected network with high social cost and a diameter in  $\Omega(\sigma(2))$ .

**Lemma 1.** *There are 2-edge-connected pairwise stable networks with  $n = \Omega(\sigma(2))$  nodes, social cost in  $\Omega(\sigma(2)n^2)$ , and diameter of at least  $\frac{\sigma(2)}{4}$ .*

The pairwise stable networks of Lemma 1, depicted in Figure 3, asymptotically reach the upper bound for the diameter of pairwise stable networks. Moreover, they allow us to prove asymptotically matching lower bounds to the PoA for the class of 2-edge-connected networks.

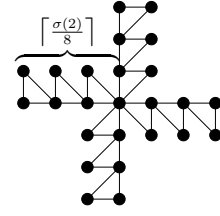


Figure 3: A high diameter pairwise stable network.

**Theorem 4.** *The PoA of SNCG is in  $\Omega\left(\frac{\sigma(n)}{n \max\{\sigma(2), n\}}\right)$ . For the class of 2-edge-connected networks the PoA is in  $\Omega(\min\{\sigma(2), n\})$ .*

We conclude this section by showing bounds to the PoS.

**Theorem 5.** *The PoS of the SNCG when  $\sigma(2) \leq 2$  or  $n$  is odd is 1. The PoS of the SNCG when  $\sigma(2) > 2$  and  $n$  is even:*

- at most  $\frac{11}{8}$  if  $\sigma(3) \geq 6$  and  $\sigma(2) \leq \frac{n}{2} - 4$ ;
- at most  $\frac{17}{12}$  if  $\sigma(2) \geq \frac{2n}{3}$ ;
- $O\left(\frac{\sigma(n)}{n \max\{\sigma(2), n\}}\right)$ , otherwise.

## Dynamics of the SNCG

So far we have considered the SNCG as a one-shot-game, i.e., we only have specified the strategy space of the agents and then focused on analyzing the equilibria of the game. In this section we focus on a more constructive sequential view of the game. As our goal is to mimick real-world social networks, we want to study the process of how such networks evolve over time. For this, we consider some initial network and then we activate the agents sequentially. An active agent will try to decrease her current cost by adding (jointly with another agent) or deleting an edge in the current network. If this process converges to a state where no agent wants to add or delete edges, then a pairwise stable network is found. Hence, such so-called improving move dynamics are a way for actually finding equilibrium states of a game. Such dynamics are guaranteed to converge if and only if the strategic game has the *finite improvement property (FIP)*, i.e., if from any strategy vector any sequence of improving strategy changes must be finite. This is equivalent to the game being a potential game (Monderer and Shapley 1996). We start with the negative result that the convergence of improving move dynamics is not guaranteed for the SNCG.

**Theorem 6.** *The SNCG does not have the FIP.*

*Proofsketch.* Let  $\sigma$  be any function such that  $2 < \sigma(2) < \frac{1}{2}\sigma(3) + 1$  and  $\sigma(3) < \frac{1}{2}\sigma(5) + 2$ .<sup>6</sup> Consider the cyclic sequence of improving moves depicted in Figure 5 in which  $G_i$  and  $G_{(i+1) \bmod 4}$  differ by exactly one edge. It is easy to show that each step of the cycle is an improving move.<sup>7</sup>  $\square$

<sup>6</sup>Observe that there are several functions  $\sigma$  that satisfy these additional constraints (for example  $\sigma(x) = \frac{3}{2}x$ ).

<sup>7</sup>A careful reader will note that each improving move is actually a best possible improving move.

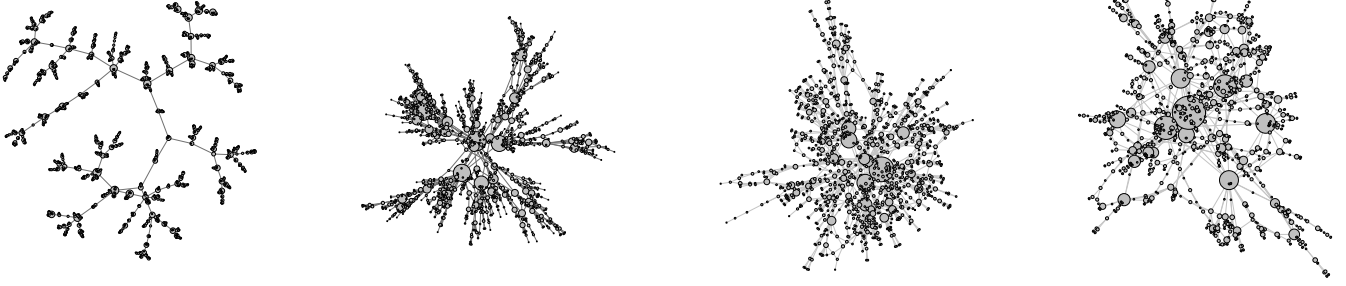


Figure 4: Snapshots of networks obtained by the iterative best move dynamic starting from a random spanning tree with  $n = 1000$  and  $\alpha = 3$ . Each plot from left to right shows the current network after 1000 steps each. The left plot shows the initial tree; the right plot shows is the final pairwise stable network. The size of the nodes is proportional to the node degrees.

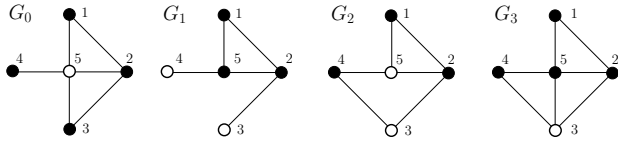


Figure 5: A cyclic sequence of improving moves. The active player or pairs of players in each step are highlighted.

The above negative result for the sequential version of the SNCG should not be overrated. In fact, when simulating the sequential process it almost always converges to a pairwise stable network. We will now discuss such simulations.

## Experimental Results

We will illustrate that starting from a sparse initial network, the sequential version of the SNCG converges to a pairwise stable network with real-world properties, like low diameter, high clustering and a power-law degree distribution. We will measure the clustering with the *average local clustering coefficient* ( $CC$ ), that is a commonly used measure in Network Science (Barabási 2016)<sup>8</sup>. Power-law degree distributions will be illustrated via log-log plots and a comparison with a perfect power-law distribution.

For all experiments<sup>9</sup> we choose  $\sigma(x) = 2 \log_2(n) \cdot x^\alpha$ , where  $n \in \mathbb{N}$  (the number of agents) and  $\alpha \in \mathbb{R}$  (the exponent) are input parameters. Clearly, this function satisfies all

<sup>8</sup>The clustering coefficient is the probability that two randomly chosen neighbors of a randomly chosen node in the network are neighbors themselves. More formally, let  $\deg(v)$  denote the degree of  $v$  in  $G$  and let  $\Delta(v)$  denote the number of triangles in  $G$  that contain  $v$  as a node. The *local clustering coefficient*  $CC(v)$  of node  $v$  in  $G$  is the probability that two randomly selected neighbors of  $v$  are neighbors, i.e.,  $CC(v) := \frac{2\Delta(v)}{\deg(v)(\deg(v)-1)}$  if  $\deg(v) \geq 2$ , and 0 otherwise. Clearly,  $0 \leq CC(v) \leq 1$ . The  $CC$  of a network  $G$  with  $n$  nodes is the average of the local clustering coefficients over all nodes  $v$ , i.e.,  $CC(G) = \frac{1}{n} \sum_{v \in V} CC(v)$ .

<sup>9</sup>The source code we used can be found at <https://github.com/melnan/distNCG.git>.

constraints we have in the definition of the game, i.e., it is convex, monotone, and  $\sigma(0) = 0$ .

Note that by Proposition 4 the upper bound for the diameter of pairwise stable networks is  $\sigma(2) + 2$  and thus we have to define  $\sigma(2)$  to be growing with  $n$  to avoid a constant diameter. Using  $\sigma(x) = 2 \log_2(n) \cdot x^\alpha$  as a proof-of-concept ensures a diameter upper bound of  $\mathcal{O}(\log n)$  that is in line with the observed diameter bounds in many real-world networks (Barabási 2016). We emphasize that also other edge cost functions with similar properties yield similar results.

In each step of our simulations one agent is activated uniformly at random and this agent then performs the best possible edge addition (jointly with the other endpoint if the respective agent agrees) or edge deletion. If no such move exists then the agent is marked, otherwise the network is updated, and all marked agents become unmarked and we repeat. The process stops when all agents are marked.

In our experiments, we always start from a sparse initial network, i.e., a cycle or a random spanning tree, to simulate an evolving social network, i.e., agents are initially connected with only very few other agents, and the number of new connections grows over time. See Figure 4 for showcase snapshots from this process.

Additional experiments starting with sparse Erdős-Rényi random networks support our intuition that the network initialization does not matter as long as the networks are sparse and the average distances are large, i.e., the resulting stable structures have the same structural properties as starting from random trees or cycles. However, for example, starting from a star network yields drastically different results. Moreover, if the initial structure is a fan graph, the algorithm stops immediately since a fan is a stable network as stated in Theorem 2. This shows that for the initial networks both sparseness and large average distances are crucial.

Figure 6 shows the box-and-whiskers plot for the average clustering coefficient of the pairwise stable networks obtained by the algorithm for  $n = 1000$  with respect to the value of the power coefficient  $\alpha$ . The upper and lower whiskers show the maximal and the minimal average clustering coefficient over 20 runs. The bottom and top of the

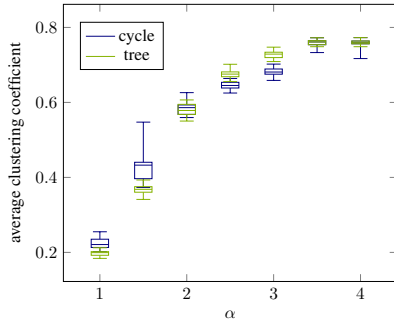


Figure 6: Average clustering coefficient of pairwise stable networks obtained by the best move dynamic for  $n = 1000$  over 20 runs with  $\sigma(x) = 18x^\alpha$ . Blue: results of the process starting from a cycle; green: starting from a random tree.

boxes are the first and the third quartiles; the middle lines are the median values. The plot explicitly shows that pairwise networks generated by the best move dynamic for a polynomial edge-cost function have a high clustering coefficient. The results indicate that the clustering coefficient correlates with the power coefficient  $\alpha$ .

Figure 7 shows a degree distribution for the resulting pairwise stable networks for  $n = 3000$ . We supplemented each plot with a plot of a perfect power-law distribution  $P(k) \sim k^{-\gamma}$ . All our experiments show that the power-law exponent  $\gamma$  is between 2 and 3, which indicates that our generated pairwise stable networks are indeed scale-free.

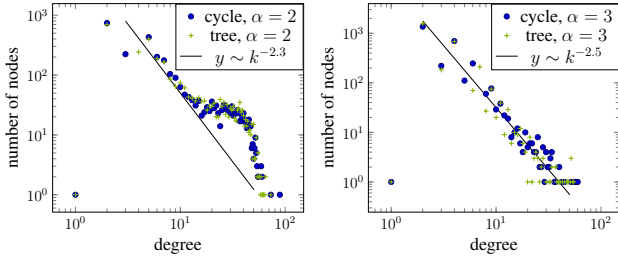


Figure 7: Log-log plot of the degree distribution of pairwise stable networks obtained by the best move dynamic for  $n = 3000$  with  $\sigma(x) = 22x^\alpha$ . Blue: results for the process starting from a cycle; green: starting from a random tree. Black: a fitted perfect power law distribution.

Finally, Figure 8 illustrates the correlation between the node degree and the local clustering coefficient of nodes with the respective degree. All plots show that the local clustering coefficient is an inverse function of a node degree. In Network Science, a local clustering following the law  $\sim k^{-1}$  is considered as an indication of the network's hierarchy that is a fundamental property of many real-world networks (Ravasz and Barabási 2003).

More plots of the degree distribution and the local clustering can be found in the full version of the paper (Bilò et al. 2020). Moreover, there we provide a comparison of our generated networks with real-world social networks.

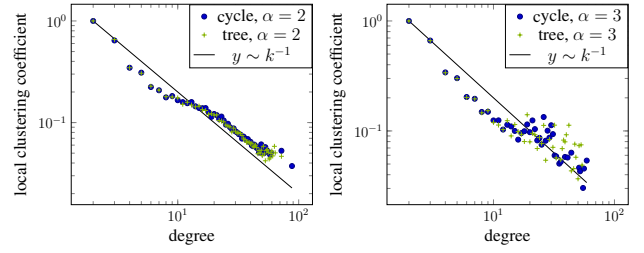


Figure 8: Log-log plot of the local clustering coefficient of nodes of a given degree in pairwise stable networks obtained by the best move dynamic for  $n = 3000$  where  $\sigma(x) = 22x^\alpha$ . Blue: results starting from a cycle; green: starting from a random tree. Black line: the function  $2/k$ .

In summary, we conclude from our proof-of-concept experiments that the best move dynamic of the SNCG generates pairwise stable networks that have very similar properties as real-world social networks.

## Conclusion

We introduced the SNCG, a promising game-theoretical model of strategic network formation where agents can bilaterally form new connections or unilaterally remove existing links aiming to maximize their centrality in the created network while at the same time to minimize the cost for maintaining edges. We emphasize that our model is based on only four simple principles: (1) agents are selfish, (2) each agent aims at increasing her centrality, (3) new connections are most likely to appear between friends of friends rather than between more remote nodes, and (4) connections are costly and can only be created by bilateral consent. All principles are motivated by modeling real-world social networks.

For this simple and stylized model for the creation of a social network by selfish agents, we provide theoretical as well as promising empirical results. On the theory side, we show that equilibrium networks of the SNCG have structural properties expected from social networks, like a high number of triangles, low diameter, and a low number of isolated 1-degree nodes. Our bounds on the PoA and the PoS show that the cost of closing a triangle essentially determines how inefficient an equilibrium network can be, compared to the social optimum. We emphasize, that all our theoretical results hold for a broad class of convex monotone edge cost functions. On the empirical side, we provide proof-of-concept results showing that the best move dynamic of the SNCG converges to equilibrium networks that share fundamental properties with real-world networks, like a power-law degree distribution, a high clustering, and a low diameter.

We see our paper as a promising step towards game-theoretic models that yield networks with all core properties of real-world networks. Future work could systematically study the influence of our model parameters on the obtained network features and on proving that the sequential network creation process indeed converges to real-world like networks with high probability.



## References

- Albers, S.; Eilts, S.; Even-Dar, E.; Mansour, Y.; and Roditty, L. 2006. On Nash Equilibria for a Network Creation Game. In *SODA'06*, 89–98.
- Albert, R.; Jeong, H.; and Barabási, A.-L. 1999. Internet: Diameter of the world-wide web. *Nature* 401(6749): 130.
- Àlvarez, C.; and Messegue, A. 2019. On the Price of Anarchy for High-Price Links. In Caragiannis, I.; Mirrokni, V. S.; and Nikolova, E., eds., *WINE'19*, 316–329.
- Anshelevich, E.; Dasgupta, A.; Kleinberg, J.; Tardos, E.; Wexler, T.; and Roughgarden, T. 2008. The price of stability for network design with fair cost allocation. *SIAM Journal on Computing* 38(4): 1602–1623.
- Bala, V.; and Goyal, S. 2000. A noncooperative model of network formation. *Econometrica* 68(5): 1181–1229.
- Barabási, A.-L. 2016. *Network science*. Cambridge University Press.
- Barabási, A.-L.; and Albert, R. 1999. Emergence of Scaling in Random Networks. *Science* 286(5439): 509–512.
- Bilò, D.; Friedrich, T.; Lenzner, P.; Lowski, S.; and Melnichenko, A. 2020. Selfish Creation of Social Networks. *arXiv preprint arXiv:2012.06203*.
- Bilò, D.; Friedrich, T.; Lenzner, P.; and Melnichenko, A. 2019. Geometric Network Creation Games. In *SPAA'19*, 323–332. ACM.
- Bilò, D.; and Lenzner, P. 2020. On the Tree Conjecture for the Network Creation Game. *Theory Comput. Syst.* 64(3): 422–443.
- Brautbar, M.; and Kearns, M. 2011. A clustering coefficient network formation game. In *SAGT'11*, 224–235. Springer.
- Bringmann, K.; Keusch, R.; and Lengler, J. 2019. Geometric inhomogeneous random graphs. *Theoretical Computer Science* 760: 35–54.
- Broder, A.; Kumar, R.; Maghoul, F.; Raghavan, P.; Rajagopalan, S.; Stata, R.; Tomkins, A.; and Wiener, J. 2000. Graph structure in the Web. *Computer Networks* 33(1): 309–320.
- Chauhan, A.; Lenzner, P.; Melnichenko, A.; and Molitor, L. 2017. Selfish Network Creation with Non-uniform Edge Cost. In *SAGT'17*, 160–172.
- Chauhan, A.; Lenzner, P.; Melnichenko, A.; and Münn, M. 2016. On Selfish Creation of Robust Networks. In *SAGT'16*, 141–152.
- Chen, Y.; Jabbari, S.; Kearns, M. J.; Khanna, S.; and Morgenstern, J. 2019. Network Formation under Random Attack and Probabilistic Spread. In *IJCAI'19*, 180–186.
- Chung, F.; and Lu, L. 2002. The average distances in random graphs with given expected degrees. *PNAS* 99(25): 15879–15882.
- Corbo, J.; and Parkes, D. 2005. The price of selfish behavior in bilateral network formation. In *PODC'05*, 99–107.
- Cord-Landwehr, A.; Mäcker, A.; and auf der Heide, F. M. 2014. Quality of Service in Network Creation Games. In *WINE'14*, 423–428. Springer.
- Demaine, E. D.; Hajiaghayi, M. T.; Mahini, H.; and Zadi-moghaddam, M. 2012. The Price of Anarchy in Network Creation Games. *ACM Trans. on Algorithms* 8(2): 13.
- Echzell, H.; Friedrich, T.; Lenzner, P.; and Melnichenko, A. 2020. Flow-Based Network Creation Games. In *IJCAI'20*, 139–145.
- Fabrikant, A.; Luthra, A.; Maneva, E.; Papadimitriou, C. H.; and Shenker, S. 2003. On a Network Creation Game. *PODC'03*, 347–351.
- Friedrich, T.; and Krohmer, A. 2015. On the diameter of hyperbolic random graphs. In *ICALP'15*, 614–625. Springer.
- Goyal, S.; Jabbari, S.; Kearns, M. J.; Khanna, S.; and Morgenstern, J. 2016. Strategic Network Formation with Attack and Immunization. In *WINE'16*, 429–443.
- Gulyás, A.; Bíró, J. J.; Kőrösi, A.; Rétvári, G.; and Krioukov, D. 2015. Navigable networks as Nash equilibria of navigation games. *Nature communications* 6: 7651.
- Jackson, M. O. 2010. *Social and economic networks*. Princeton university press.
- Jackson, M. O.; and Rogers, B. W. 2005. The economics of small worlds. *Journal of the European Economic Association* 3(2-3): 617–627.
- Jackson, M. O.; and Wolinsky, A. 1996. A strategic model of social and economic networks. *Journal of economic theory* 71(1): 44–74.
- Kleinberg, J. 2000. The Small-world Phenomenon: An Algorithmic Perspective. In *STOC'00*, STOC '00, 163–170. New York, NY, USA: ACM. ISBN 1-58113-184-4.
- Kouroupas, G.; Markakis, E.; Papadimitriou, C.; Rigas, V.; and Sideri, M. 2015. The web graph as an equilibrium. In *SAGT'15*, 203–215. Springer.
- Koutsoupias, E.; and Papadimitriou, C. 1999. Worst-case Equilibria. In *STACS'99*, 404–413.
- Krioukov, D.; Papadopoulos, F.; Kitsak, M.; Vahdat, A.; and Boguñá, M. 2010. Hyperbolic geometry of complex networks. *Phys. Rev. E* 82: 036106.
- Leskovec, J.; Kleinberg, J.; and Faloutsos, C. 2005. Graphs over time: densification laws, shrinking diameters and possible explanations. In *SIGKDD'05*, 177–187. ACM.
- Mamagishvili, A.; Mihalák, M.; and Müller, D. 2013. Tree Nash Equilibria in the Network Creation Game. In *WAW'13*, volume 8305 of *LNCS*, 118–129.
- Meirom, E. A.; Mannor, S.; and Orda, A. 2014. Network formation games with heterogeneous players and the internet structure. In *EC'14*, 735–752. ACM.
- Meirom, E. A.; Mannor, S.; and Orda, A. 2015. Formation games of reliable networks. *INFOCOM '15*, 1760–1768. IEEE.



Mihalák, M.; and Schlegel, J. C. 2010. The Price of Anarchy in Network Creation Games is (mostly) constant. *SAGT'10*, 276–287.

Monderer, D.; and Shapley, L. S. 1996. Potential Games. *Games and Economic Behavior* 14(1): 124 – 143. ISSN 0899-8256.

Newman, M.; Barabasi, A.-L.; and Watts, D. J. 2011. *The structure and dynamics of networks*. Princeton University Press.

Papadimitriou, C. H. 2001. Algorithms, games, and the internet. In *STOC'01*, 749–753.

Ravasz, E.; and Barabási, A.-L. 2003. Hierarchical organization in complex networks. *Physical review E* 67(2): 026112.