Games and Economic Behavior 75 (2012) 694–713

Contents lists available at SciVerse ScienceDirect

Games and Economic Behavior

www.elsevier.com/locate/geb

The efficiency and stability of R&D networks

Michael D. König [a,b,*], Stefano Battiston [b], Mauro Napoletano [c,b], Frank Schweitzer [b]

[a] SIEPR and Department of Economics, Stanford University, 579 Serra Mall, CA 94305-6072, United States
[b] Chair of Systems Design, Department of Management, Technology and Economics, ETH Zurich, Kreuzplatz 5, CH-8032 Zurich, Switzerland
[c] OFCE-Sciences Po and SKEMA Business School, 250 rue Albert Einstein, 06560 Valbonne, France

ARTICLE INFO

Article history:
Received 30 June 2009
Available online 14 January 2012

JEL classification:
D85
L24
O33

Keywords:
R&D networks
Knowledge recombination
Indirect spillovers
Network efficiency
Network formation

ABSTRACT

We investigate the efficiency and stability of R&D networks in a model with network-dependent indirect spillovers. We show that the efficient network structure critically depends on the marginal cost of R&D collaborations. When the marginal cost is low, the complete graph is efficient, while high marginal costs imply that the efficient network is asymmetric and has a nested structure. Regarding the stability of network structures, we show the existence of both symmetric and asymmetric equilibria. The efficient network is stable for small industry size and small cost. In contrast, for large industry size, there is a wide region of cost in which the efficient network is not stable. This implies a divergence between efficiency and stability in large industries.

1. Introduction

This paper investigates the efficiency and evolution of networks of firms engaged in costly R&D collaborations. Collaborations among firms in inventive activities have become very common, especially in those industries characterized by rapid technological change (e.g. the pharmaceutical, chemical and computer industries, see Hagedoorn, 2002; Powell et al., 2005). R&D collaborations allow firms to directly combine the knowledge, skills and physical assets needed to innovate. In addition, they provide access to indirect spillovers, by serving as conduits through which knowledge and information can spread across firms (see Ahuja, 2000; Powell et al., 2005). As a consequence, collaborations do not occur between two isolated actors, but rather involve firms that are already part of a network of partnerships with other firms. The increasing importance of such a phenomenon has also spurred economic research on the structural features of the network of R&D collaborations, and on their impact on industry performance. On the first issue, empirical studies have shown that real-world networks are typically asymmetric, i.e. featuring at the same time firms having many collaborations and others with few links (see e.g. Powell et al., 2005). Regarding efficiency, the debate has focused on whether densely interconnected networks are more conducive to industry performance than an asymmetric network featuring "structural holes", i.e. displaying the presence of hubs indirectly connecting many firms which have no direct link across them (Burt, 1992; Ahuja, 2000; Powell et al., 1996; Westbrock, 2010).

Motivated by the above research stream we study a model of strategic network formation in which firms innovate by recombining knowledge with that of their R&D partners in the industry. We formalize this idea by assuming that (i) the number of innovations is proportional to the stock of knowledge of the firm, and that (ii) a firm's knowledge growth is a

* Corresponding author.
E-mail address: mdkoenig@stanford.edu (M.D. König).

0899-8256/$ – see front matter
doi:10.1016/j.geb.2011.12.007

linear combination of the idiosyncratic knowledge stock of the firm and the knowledge of its R&D partners. In the model, firm's discounted profits are a function of the number of innovations and include the costs of R&D collaborations. In this framework, if firms' discount rate is small then the gross payoff from collaborations is proportional to the largest eigenvalue of the adjacency matrix associated with the connected component to which the firm belongs. The largest eigenvalue is related to the number of all walks connecting firms in a given component (see Cvetkovic et al., 1997, pp. 24). This has several implications. First, as the largest eigenvalue is the same for all firms in the same component, the formation/deletion of a collaboration by a firm has a strong non-rival external effect on all its direct and indirect neighbors. Second, both individual and industry profits induce a trade-off between the need of sustaining walks via direct connections and the costs of such direct connections. The strength of this trade-off is captured by the value of marginal collaboration costs. Third, the decision whether a link is formed or deleted depends on the change in eigenvalue, which varies with the position of the two firms involved in the collaboration. In addition, the change in eigenvalue is inversely related to the number of firms in the industry, while marginal costs stay constant with industry size. This implies that firms have stronger incentives to form collaborations in small industries than in large ones.

The model builds on previous work of the authors (König et al., 2008, and in particular König et al., 2011). In König et al. (2011), we studied the structural features of equilibrium networks arising from the model and their relations to observed properties of empirical networks. In this paper, we focus instead on the analysis of network efficiency and of the possible tension between equilibrium and efficiency. We first characterize the topology of the efficient structure for varying marginal cost of collaborations. When the marginal costs of collaboration are low, the trade-off between number of walks and costs of direct connections is loose, and we show that the complete network is efficient. For intermediate values of marginal collaboration costs, efficient graphs are still connected but it becomes efficient to save on the number of direct collaborations and to create a high number of walks by concentrating connections among a small number of firms. More precisely, efficient graphs are nested split graphs, i.e. the neighborhood of each node is contained in the neighborhood of the next higher degree nodes. This implies a strong hierarchical degree structure of the network. As costs of collaboration become high, asymmetric efficient graphs can become disconnected and then empty. Next, we study the emergence of pairwise stable structures (cf. Jackson and Wolinsky, 1996). The complete graph is stable if marginal costs of collaboration are low and industry size is small. In large industries, we find that the set of disconnected cliques of homogeneous size and the star are stable. Interestingly, both structures can be stable for the same values of industry size and collaboration costs. In addition, the size of stable cliques decreases with marginal collaboration costs. The empty graph is stable for very large marginal costs and any industry size. Finally, we study the relationship between stability and efficiency and we show that efficiency is reached in industries of small size and low cost of collaboration. As industry size grows, however, firms have lower incentives to form collaborations, and a divergence between stability and efficiency emerges.

R&D networks have been theoretically studied in Goyal and Moraga-Gonzalez (2001), Goyal and Joshi (2003), Westbrock (2010). As in the two latter works, we assume that the effort in collaborations is exogenous and we focus on the trade-off between the benefits from collaborations and the costs of maintaining them. However, differently from these contributions, we study efficiency and stability of R&D networks in the case of indirect spillovers that depend on the structure of the network to which a firm belongs. This has several implications for the results. First, Goyal and Joshi (2003), Westbrock (2010) show that asymmetric networks can be efficient under direct and industry-wide spillovers. Our results therefore extend the efficiency of asymmetric networks to the situation in which network-dependent indirect spillovers are also present. In particular, our efficient graphs, when connected, are strictly contained in the class of interlinked stars, and when disconnected contain only one non-trivial component as in the dominant architecture. Furthermore, both Westbrock (2010) and Goyal and Joshi (2003) show that stable structures are typically asymmetric at significant values of linking costs. We obtain a similar result (the stability of the star). However, differently from both papers, we find that also perfectly symmetric network structures (the set of disconnected cliques) can be stable for significant values of the cost of link formation, even in the same region of parameters wherein asymmetric structures are stable. Finally, like in Westbrock's paper, we find that the properties of stable graphs can significantly diverge from those of efficient graphs in large industries. Compared to efficient graphs, stable graphs in our model are "under-connected" with respect to what is socially desirable.

R&D collaborations with network-dependent indirect spillovers have already been investigated in Deroian (2008). However, we depart from this model in two respects: (i) we perform the analysis of network stability and efficiency also for non-negligible costs of link formation, and (ii) indirect spillovers are not constant within-components, but endogenously change with the topology of the component to which the firm belongs.

Finally, we discuss the relations between our model and the contributions in the network formation literature that have focused on the trade-off between the costs of connections and the direct and indirect benefits from the network (see Vega-Redondo, 2007; Goyal, 2009; Jackson, 2008, for recent surveys). Our model shares many similarities with the contributions of Jackson and Wolinsky (1996), Bala and Goyal (2000), Goyal and Joshi (2006), Ballester et al. (2006). We contribute to this literature by studying both network efficiency and evolution in a framework with positive network externalities, and in which the gross payoff the agent receives from the network positively depends on all walks the agent can directly and indirectly access by establishing a connection. Incorporating all walks in agents' payoffs has also implications for the results obtained. First, efficient graphs have in general a more complex structure than in the models cited above (e.g. they are not minimally connected, cf. Bala and Goyal, 2000). Second, we show that disconnected graphs (the cliques) are stable in the presence of positive network externalities, whereas this result has so far been typically derived in models with negative externalities (see e.g. Jackson and Wolinsky, 1996; Goyal and Joshi, 2006).

The paper is organized as follows. Section 2 contains the description of the model. Section 3 is devoted to the analysis of the efficiency of R&D network structures. Stable networks are analyzed in Section 4. The relation between efficiency and stability is studied in Section 5. Finally, Section 6 concludes. All proofs can be found in Appendix A.

2. The model

We consider a two-stage process of link formation and knowledge accumulation (see e.g. Goyal and Joshi, 2003; Westbrock, 2010). In the first stage, firms form pairwise R&D collaborations with other firms in the same industry, making possible the growth of knowledge within each firm. In the second stage, firms introduce innovations in the market by using the knowledge accumulated through collaborations. We first define the network of R&D collaborations. Next, we characterize the payoffs from R&D collaborations. Finally, we briefly discuss the relations between our model and the literature on strategic network formation.

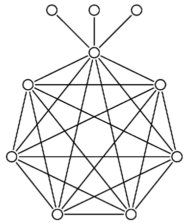
2.1. The network

Consider an industry populated by $n > 2$ firms. The *network* or *graph* G is the pair (N, E) consisting of the set of nodes $N(G) = \{1, \dots, n\}$ representing the population of firms, and a set of *edges* $E(G)$, or *links*, representing R&D collaborations among the firms (for simplicity we may just write N and E where it is clear which network G the sets refer to).[1] The number of nodes is $n = |N|$ and the number of edges $m = |E|$. An edge $ij \in E$, represents the existence of an R&D collaboration between firm i and j, which are said to be *adjacent*. The *neighborhood* of a node i is the set $N_i = \{j \in N\colon\ ij \in E\}$. The *degree* of a node i in G, written by d_i, is the number of edges incident to i. Clearly, $d_i = |N_i|$. The maximum degree is $\triangle(G)$ and the minimum degree is $\delta(G)$. A *symmetric graph* or *regular graph* is a graph where each node has the same degree. A *walk* $W(i_1, i_k)$ connecting firms i_1 and i_k is a sequence of firms $(i_1, i_2, \dots, i_k)$ such that $i_1i_2, i_2i_3, \dots, i_{k-1}i_k \in E$. A walk is *closed* if the first and last firms in the sequence are the same, and *open* if they are different. The length of the walk is given by the number of edges it contains, i.e. any walk $W(i_1, i_k)$ has length $k - 1$. A *path* is a walk in which no firm is visited twice. A closed path encompassing n nodes is a *cycle*, denoted by C_n. A *connected component* in G is a maximal set of firms such that there exists a path between any two of them. We will say that two components are *disconnected* if there is no path between them. A *connected graph* is a graph consisting of only one connected component.

Let $\mathbf{A}(G)$ be the symmetric $n \times n$ *adjacency matrix* of the R&D network G. The element $a_{ij} \in \{0, 1\}$ indicates if there exists a link between nodes i and j such that $a_{ij} = 1$ if $ij \in E$ and $a_{ij} = 0$ if $ij \notin E$. The *eigenvalues* of the adjacency matrix $\mathbf{A}$ are the numbers λ such that $\mathbf{Av} = \lambda\mathbf{v}$ has a nonzero solution vector $\mathbf{v}$, which is an *eigenvector* associated with λ. λ_{PF} denotes the *largest real eigenvalue* of $\mathbf{A}$ (the *Perron–Frobenius eigenvalue*, cf. Seneta, 2006; Horn and Johnson, 1990), i.e. all eigenvalues λ of $\mathbf{A}(G)$ satisfy $|\lambda| \leqslant \lambda_{\mathrm{PF}}$ and there exists an associated non-negative eigenvector $\mathbf{v} \geqslant 0$ such that $\mathbf{Av} = \lambda_{\mathrm{PF}}\mathbf{v}$. For a connected graph G, $\mathbf{v}$ is strictly positive.

A *subgraph* of G is a pair $G' = (N', E')$ such that $N' \subseteq N$, $E' \subseteq E$, and $\forall ij \in E'$ it holds $i, j \in N'$. We write $G' \subseteq G$ if G' is a subgraph of G. A *complete graph* K_n is a graph in which all n nodes are pairwise adjacent. The graph in which no pair of nodes is adjacent is the empty graph $\bar{K}_n$. A *clique* $K_{n'}$, $n' \leqslant n$, is a complete subgraph of the network G. A set of disconnected cliques or *multiple cliques* consists of $d \geqslant 2$ disconnected cliques of the same size k, $K_k^1, \dots, K_k^d$. The *dominant-group architecture* $D_{n,k}$ consists of a clique of k nodes and $n - k$ isolated nodes. Consider a partition of the nodes of a connected graph G in classes of increasing degree, $\{D_1, D_2, \dots, D_k\}$. Assume that $k \geqslant 2$. Then G is an *interlinked star*, if the neighborhood of each node in D_1 is D_k and each node in D_k is connected to all other $n - 1$ nodes (Goyal and Joshi, 2003). A *nested split graph*[2] (see Aouchiche et al. (2008) and Cvetkovic et al. (1997, pp. 60–74)) is a graph with a *stepwise* adjacency matrix $\mathbf{A}$, i.e. its elements a_{ij} satisfy the condition that if $i < j$ and $a_{ij} = 1$ then $a_{hk} = 1$ whenever $h < k \leqslant j$ and $h \leqslant i$. In a stepwise matrix, if there exists an element equal to one, $a_{ij} = 1$, then also the element in the row above in the matrix is one, $a_{i,j-1} = 1$, and the element in the column to the left in the matrix is one, $a_{i-1,j} = 1$. Consequently, all the preceding elements to the left and above are one. A *nested star*,[3] denoted by $F_{n,d}$ (Bell, 1991), is the graph obtained from the complete graph K_d with d nodes and a subset of $n - d$ disconnected nodes, by adding $n - d$ links connecting one node in K_d to each of the $n - d$ disconnected nodes. Notice, that the nested star has a stepwise adjacency matrix, and therefore is a special case of connected nested split graph. Fig. 1 shows an example of connected nested split graph (the nested star $F_{10,7}$) with the associated adjacency matrix. Finally, notice that the complete graph and the spanning star are particular cases of connected nested split (and nested star) graphs. Further properties of nested split graphs are discussed in Section 3.

[1] We consider undirected graphs only.
[2] Nested split graphs are also known as "threshold graphs" (cf. Mahadev and Peled, 1995).
[3] The name is introduced here in analogy to the notion of interlinked star.



$$\mathbf{A} = \begin{pmatrix} 0&1&1&1&1&1&1&1&1&1\\ 1&0&1&1&1&1&1&0&0&0\\ 1&1&0&1&1&1&1&0&0&0\\ 1&1&1&0&1&1&1&0&0&0\\ 1&1&1&1&0&1&1&0&0&0\\ 1&1&1&1&1&0&1&0&0&0\\ 1&1&1&1&1&1&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0\\ 1&0&0&0&0&0&0&0&0&0 \end{pmatrix}$$

Fig. 1. An example of a connected nested split graph, the nested star $F_{10,7}$, (left) and its associated adjacency matrix (right).

2.2. Payoffs from collaborations

The modeling of the payoffs from R&D collaborations is similar to the one[4] in König et al. (2011). Collaborations allow the growth of knowledge within the firm, via the recombination of the existing knowledge of the firm and of its partners (see Weitzman, 1998; Kogut and Zander, 1992). Formally, new knowledge within firm $i \in N$ is generated in continuous time according to[5]

$$\dot{x}_i(t) = \gamma x_i(t) + \beta \sum_{j=1}^{n} a_{ij} x_j(t), \quad \gamma > 0,\ \beta > 0, \tag{1}$$

where γ and β measure the response of knowledge growth to, respectively, variations in internal and external knowledge.[6] Without any loss of generality, in what follows we set $\beta = 1$. The growth of knowledge of firm i in Eq. (1) depends both on spillovers emanating from the direct neighbors of i, as well as on spillovers from indirect neighbors which affect the stock of knowledge of direct neighbors. It follows that the topology of the whole network of R&D collaborations (including all direct and indirect paths along which knowledge can flow between firms) influences knowledge dynamics within the firm. The presence of network-dependent indirect spillovers in the analysis is an important feature of our model, that distinguishes it from previous R&D networks models (e.g. Goyal and Joshi, 2003; Westbrock, 2010). Indeed, in these models the focus has almost exclusively been on direct or industry-wide spillovers.[7]

Furthermore, in line with previous works in the literature (see Goyal and Moraga-Gonzalez, 2001; Goyal and Joshi, 2003), we assume that returns $R(x_i(t))$ from innovations at time t are an increasing and concave function of the knowledge stock $x_i(t)$ of firm i according to[8]

$$R\big(x_i(t)\big) = V \ln x_i(t), \quad V > 0. \tag{2}$$

Each collaboration involves an increasing cost per unit of time $\tilde{c}_0 + \tilde{c}_1 t$.[9] If firm i engages in d_i collaborations then its total cost per unit of time is equal to $d_i(\tilde{c}_0 + \tilde{c}_1 t)$. Given the network G, a firm forms or severs a collaboration by evaluating the future discounted profits from the collaboration. Without loss of generality, we set to zero the time at which a link is formed or severed. Total discounted profits of firm i from collaborations with d_i other firms are equal to

$$\tilde{\pi}_i(G_i, \tilde{\mathbf{c}}, \rho) = \int_0^\infty \big(V \ln x_i(t) - d_i(\tilde{c}_0 + \tilde{c}_1 t)\big) e^{-\rho t}\, dt, \tag{3}$$

where $\rho > 0$ is the discount rate and $G_i \subseteq G$ is the connected component of firm i.

To obtain further insights into the structural determinants of the payoff function in Eq. (3), we study the asymptotic properties of knowledge growth in Eq. (1). König et al. (2011) show that, for a given network G of collaborations, the

[4] König et al. (2011) use a timing scheme for agents' decision that differs from the one in this paper. However, this difference has no impact on the properties of agents' payoffs discussed below.

[5] A similar knowledge growth function can also be found in Spence (1984). In addition, note that Eq. (1) implies that knowledge growth is "cumulative", a property often emphasized in empirical innovation studies (e.g. Dosi, 1988).

[6] In vector-matrix notation Eq. (1) reads $\dot{\mathbf{x}}(t) = (\mathbf{A}(G) + \gamma \mathbf{I})\mathbf{x}(t)$. Note also that for positive initial values, $\mathbf{x}(0) > 0$, knowledge stocks remain positive over time, $\dot{\mathbf{x}}(t) > 0$ as well as $\mathbf{x}(t) > 0$, $\forall t \geqslant 0$.

[7] An exception is the R&D networks model of Deroian (2008). In this work, however, the analysis is limited to the case of zero costs of link formation. Moreover, the characteristics of network benefits in our model have significant differences with those of Deroian's model. See Section 2.3 for more discussion.

[8] The logarithmic form of the return function in Eq. (2) is both increasing and concave in the knowledge level. In addition, it allows in a simple way the convergence of discounted revenues from innovation in Eq. (3).

[9] The assumption that collaboration costs increase over time, avoids that total discounted costs become negligible compared to revenues in Eq. (3). In this model the rationale of R&D collaboration lies in the possibility of absorbing external knowledge. Thus, collaboration costs mainly account for the R&D investments needed to acquire this external knowledge. Empirical works, such as Cohen and Levinthal (1989) and Veugelers (1997), show that the possibility of acquiring knowledge from external sources is directly to internal R&D investments of the firms. Moreover, the fact that R&D costs display a positive trend is also documented by empirical evidence (see e.g. Jones, 1995; Copeland and Fixler, 2009; DiMasi et al., 2003).

knowledge dynamics introduced in Eq. (1) converges to a steady state in which a firm's knowledge growth rate depends on the topology of the connected component $G_i \subseteq G$ to which the firm i belongs, through the largest eigenvalue $\lambda_{\text{PF}}(G_i)$ (see Proposition 1 in König et al., 2011). More formally,

$$\lim_{t\to\infty} \frac{\dot{x}_i(t)}{x_i(t)} = \lambda_{\text{PF}}(G_i) + \gamma. \tag{4}$$

By exploiting the above result, and by assuming that firms are weakly discounting future revenues (i.e. as $\rho \to 0$), total discounted payoffs from innovations can be approximated with the following expression (see Proposition 2 in König et al., 2011):

$$\tilde{\pi}_i(G_i, \tilde{\mathbf{c}}, \rho) \approx \frac{V(\lambda_{\text{PF}}(G_i) + \gamma)}{\rho^2} - \frac{\tilde{c}_1 d_i}{\rho^2}. \tag{5}$$

Since the discount rate is identical for all firms, we apply an affine transformation to obtain the final expression for the payoff of firm i from R&D collaborations

$$\pi_i(G_i, c, \rho) = \lambda_{\text{PF}}(G_i) - c d_i, \tag{6}$$

where $c = \tilde{c}_1/V$. Eq. (6) implies that the payoff from collaborations depends on the topology of the network. In particular, the largest eigenvalue associated with a component, $\lambda_{\text{PF}}(G_i)$, coincides with the growth rate in the number of walks of length k in a component,[10] when the length is increased by one (see Cvetkovic et al., 1997, p. 24). On the other hand, the payoff decreases with the degree d_i of the firm. Therefore, it is best for a firm to reach the other firms through many walks but to have not too many links to pay for.

2.3. The largest real eigenvalue, indirect spillovers and network formation

Eq. (6) implies that the revenues from collaborations are the same for all firms belonging to the same connected component. However, marginal revenues from creating or terminating a collaboration display both between- and within-component variability.[11] Indeed, the change in the largest real eigenvalue varies with the position of the firm that creates or removes a link. For instance, in a star of n firms, the marginal revenue that a disconnected firm gains from creating a link to the hub is typically larger than the marginal revenue that two peripheral firms gain from creating a link between them (see the proof of Proposition 6 in König et al., 2011). The following lemma (see Cvetkovic et al., 1997, pp. 133, 50; and Cvetkovic and Rowlinson, 1990) describes the relationship existing between the largest real eigenvalue λ_{PF} and the addition/removal of a link in the network.[12]

Lemma 1. *Denote $G' = (N', E')$ the graph obtained from the graph $G = (N, E)$ by the addition or removal of an edge ij. Then the following properties hold*:

1. $\lambda_{\text{PF}}(G') \geqslant \lambda_{\text{PF}}(G)$ *if* $ij \notin E$ *and* $\lambda_{\text{PF}}(G') \leqslant \lambda_{\text{PF}}(G)$ *if* $ij \in E$, *with inequalities being strict if the graph is connected.*
2. $\lambda_{\text{PF}}(G) \leqslant \lambda_{\text{PF}}(K_n) = n - 1$.
3. $|\lambda_{\text{PF}}(G') - \lambda_{\text{PF}}(G)| \leqslant 1$.

Item (i) in Lemma 1 implies that the gross payoff function in our model displays monotonicity to addition/removal of links (cf. Goyal and Joshi, 2006), and strong monotonicity when the graph is connected. Moreover, since the gross payoff from collaboration $\lambda_{\text{PF}}(G_i)$ is the same for all firms in the same component G_i of firm i, the same item implies that the creation (deletion) of a link by the firm has a positive (negative) non-rival external effect on all its direct and indirect neighbors in the component. Furthermore, Lemma 1 states that the gross payoff is bounded from above by the largest eigenvalue associated with the complete graph K_n. In addition, the marginal gross payoff is also bounded, since its value must be less than one.

Besides the properties mentioned in Lemma 1, all network architectures that we analyze in this paper are characterized by an inverse relationship between the marginal gross payoff and the size of the connected component (see also Sections 4 and 5). Since marginal cost does not vary with network size, the foregoing property implies that firms will have stronger incentives to form collaborations in small industries than in large ones.[13]

[10] Interestingly, empirical studies on R&D networks (e.g. Powell et al., 2005) support the idea that firms establish R&D collaborations in a way that increases the number of walks in the network.

[11] Furthermore, notice that although total discounted profits from innovation are independent from the position of each firm i in the network, the instantaneous returns $R(x_i(t))$ depend on the network position of the firm via the stock of knowledge $x_i(t)$.

[12] Note that N in Lemma 1 is not necessarily the same as N'.

[13] There are some other useful bounds on λ_{PF} besides those discussed in Lemma 1 (see e.g. Cvetkovic and Rowlinson, 1990, for a more complete account). For a graph G with m links and n nodes it holds that $\lambda_{\text{PF}}(G) \leqslant \sqrt{2m(1-\frac{1}{n})}$, with equality if G is the complete graph K_n. If, in addition, G is connected then $\lambda_{\text{PF}}(G) \leqslant \sqrt{2m - n + 1}$, with equality if G is the star $K_{1,n-1}$ or the complete graph K_n. Finally, let $\delta(G)$, $\bar{d}(G)$, $\Delta(G)$ be the minimum, mean and maximum degree in the connected graph G, respectively. Then $\delta(G) \leqslant \bar{d}(G) \leqslant \lambda_{\text{PF}}(G) \leqslant \Delta(G)$.

The payoff function in Eq. (6) can be compared to other similar functions in the network formation literature. First, in the "connections" model of Jackson and Wolinsky (1996) the benefit of an agent depends on the shortest paths to other agents in the network. Our model instead takes into account all possible walks from agent i to the other agents in the same connected component. This is in line with the idea of knowledge recombination, in which firms improve their knowledge by recombining the knowledge of different firms (e.g. Weitzman, 1998). In such a framework, spillovers emanating from the same source are worth differently to a firm if they reach it along different paths. It is also important to stress that such an argument requires that technological and spatial distance are not important for spillovers diffusion. In contrast, shortest paths are better suited for environments in which either technological or spatial distance is important, and wherein firms get utility from actively searching for the closest neighbors (see Carayol and Roux, 2009; Deroian, 2008, for models wherein distance across firms plays a role).

Second, Bala and Goyal (2000) and Deroian (2008) introduce models in which the payoff function depends either on the number of nodes or on the total number of links of the connected component the agent belongs to. In contrast, in our model the payoff is affected, through the largest eigenvalue, by the whole topology of the component.[14]

Walks across the network play a key role also in the model of Ballester et al. (2006). There, the equilibrium effort of every agent is proportional to her eigenvector centrality, while in our model, the gross payoff of agent i is proportional to the largest eigenvalue associated with her connected component G_i. The eigenvector centrality of an agent essentially counts the number of walks departing from that agent, discounting longer walks by means of a decaying factor. Similarly, the largest eigenvalue also grows with the number of walks. However, the largest eigenvalue is the same for all the agents belonging to a connected component while the eigenvector centrality is not.

3. Efficiency

In the model presented in the previous section, firms face a trade-off between increasing the number of walks in the network and the costs of creating such walks via direct connections with other firms. The linearity of the firm's payoff function (cf. Eq. (6)) implies that such a trade-off also characterizes industry profits, and in this section we investigate how this trade-off can be managed to yield the most efficient outcome for the industry. We start by defining industry welfare as the sum of firms' profits (cf. Jackson and Wolinsky, 1996)

$$\Pi(G,c)=\sum_{i=1}^{n}\pi_i(G_i)=\sum_{i=1}^{n}\big(\lambda_{\mathrm{PF}}(G_i)-cd_i\big)=\sum_{i=1}^{n}\lambda_{\mathrm{PF}}(G_i)-2mc. \tag{7}$$

We are interested in finding the network structures that maximize Eq. (7) for a given level of marginal cost c. This is because, the level of marginal cost c captures how strong is the trade-off between walks and cost of direct connections in the model. More formally, let $\mathcal{G}(n)$ denote the set of all graphs with n nodes. For a given value of cost c, the efficient graph is defined as $G^*=\arg\max_{G\in\mathcal{G}(n)}\Pi(G,c)$.

In general, the determination of the efficient graph is a hard combinatorial optimization problem. A first result that we derive concerns connectedness. For $c\in[0,1)$, a graph is efficient only if it is connected, while for $c\in[1,\infty)$ it can be efficient to have a single connected component and isolated nodes.

Lemma 2. *Consider a graph G consisting of two disconnected components G_1 and G_2, with n_1, n_2 nodes, m_1, m_2 edges, eigenvalues $\lambda_{\mathrm{PF}}(G_1)$, $\lambda_{\mathrm{PF}}(G_2)$ and total profits $\Pi(G_1)=n_1\lambda_{\mathrm{PF}}(G_1)-2m_1c$, $\Pi(G_2)=n_2\lambda_{\mathrm{PF}}(G_2)-2m_2c$. Then there exists a graph G' with $n=n_1+n_2$ nodes and aggregate profits satisfying $\Pi(G')>\Pi(G)=\Pi(G_1)+\Pi(G_2)$ such that*

(i) *if $c\in[0,1)$ then G' is connected,*
(ii) *if $c\in[1,\infty)$ then G' consists of at most one non-singleton connected component.*

Thus, in order to guarantee efficiency when $c\in[0,1)$, each firm must have (direct or indirect) access to the knowledge of all other firms in the industry. For $c>1$ the efficient graph can be disconnected. However, in this case the efficient graph features only one connected component rather than several disconnected structures.

Since in $c\in[0,1)$ the efficient graph is connected, Eq. (7) for total profits boils down to

$$\Pi(G,c)=n\lambda_{\mathrm{PF}}(G)-2mc. \tag{8}$$

Note that isolated nodes neither contribute to the eigenvalue nor to the linking cost, so that for $c>1$, the same equation holds with G, n and m referring to the only non-trivial connected component of the graph. Thus, for any given values of n and m, the efficient graph G^* always belongs to the class of connected graphs that maximize $\lambda_{\mathrm{PF}}(G)$. This class is known to

[14] In this regard, our model also generalizes the notion of local spillover games (with positive spillovers) studied in Goyal and Joshi (2006). In local spillover games the marginal payoff to agent i from a link with agent j depends only on the number of links of i and j (and it is independent of the degree of other agents $k\neq i,j$). In our model instead, the marginal gross payoff for i from linking to j is not just a function of the degree of the two agents, but it depends (positively) on all indirect partners that can be reached through that connection.

coincide with a special class of graphs, called "nested split graphs" (cf. Brualdi and Solheid, 1986; see also Section 2.1 for the definition). The following proposition characterizes efficient structures for marginal cost of collaboration $c \in [0, \infty)$.[15,16]

Proposition 1. *Let G^* be the efficient graph for a given number $n > 3$ of firms and cost of collaboration $c \geqslant 0$.*

(i) *If $c \in [0, \frac{n}{2n-1}]$ ($\sim [0, 0.5]$ for large n) then the unique efficient network G^* is the complete graph.*
(ii) *If $c \in [\frac{n}{2n-3}, 1]$ ($\sim [0.5, 1]$ for large n) then the efficient network G^* is a connected nested split graph which is not the complete network.*
(iii) *If $c \in [1, \frac{n}{2\sqrt{n-1}})$ then the efficient network G^* is not empty and has one connected component which is a nested split graph.*
(iv) *If $c \in [\frac{n}{2\sqrt{n-1}}, \sqrt{\frac{n(n-1)}{2}}]$ then the efficient network G^* has at most one non-singleton component which is a nested split graph.*
(v) *If $c > \sqrt{\frac{n(n-1)}{2}}$ ($\sim n$ for large n) then the unique efficient network G^* is the empty graph.*

Notice that the complete graph and the empty graph are special cases of nested split graphs. However, the complete graph is the only symmetric graph in this class as all other graphs contain at least one maximally connected node (with degree $n-1$) and some nodes with degree strictly smaller. The most relevant property of nested split graphs is that they have a nested neighborhood structure. Consider a partition of nodes in the network G by increasing degree, $\{D_1, D_2, \ldots, D_k\}$, where the nodes in D_1 have the minimum degree $\delta(G)$, the nodes in D_k have maximum degree $\Delta(G)$ and $d_i < d_j$ for all $i \in D_l$, $j \in D_{l+1}$, $1 \leqslant l \leqslant k-1$. Then, in a nested split graph, the neighborhood of node i is contained in the neighborhood of node j, i.e. $N_i \subset N_j$, if $d_i < d_j$ (Mahadev and Peled, 1995, Theorem 1.2.4, p. 10). Moreover, if G is a connected nested split graph then it also holds that $\Delta(G) = n-1$. From the above description it follows that the class of asymmetric connected nested split graphs is strictly contained in the class of interlinked stars introduced in Goyal and Joshi (2003).[17]

Furthermore, the nested neighborhood structure of an asymmetric efficient graph can be thought of as a generalization of the concept of core–periphery architecture (see among others Borgatti and Everett, 2000; Hojman and Szeidl, 2008; Bramoullé, 2007; Galeotti and Goyal, 2009; Persitz, 2009). Usually, core–periphery architectures describe networks with one group of densely connected agents (the core) and another group of agents (the periphery) connected to the core, who are only sparsely connected among themselves. The difference is that graphs with a nested neighborhood structure can feature several densely connected groups with nodes of increasing degree. In turn, each group is connected to the group of higher degree nodes.

It is also possible to provide some constraints on the number m^* of links in the efficient graph G^*, which follow already from our previous analysis. First, if $c \in [0, 0.5]$ then $m^* = \binom{n}{2}$, because the efficient graph is the complete graph K_n. Second, if $c \in [0, 1)$ then $m^* \geqslant n-1$, because the efficient graph is connected in that region of cost. Third, Proposition 1 implies that, if $c > \sqrt{\frac{n(n-1)}{2}}$ then $m^* = 0$. Beyond these preliminary observations, the next proposition derives upper and lower bounds for the number of links m^* in the efficient network G^* as a function of cost c.[18]

Proposition 2. *Let m^* denote the number of links in the efficient graph G^* and denote by $c_1^* = \frac{n+2\sqrt{2n^2(n-1)}}{8n-9} < c_2^* = \sqrt{\frac{n(n-1)}{2}}$.*

(i) *If $c \in [0, \frac{n}{2n-1}]$ ($\sim [0, 0.5]$ for large n), then $m^* = \binom{n}{2}$.*
(ii) *If $c \in (\frac{n}{2n-1}, c_1^*)$, then $\exists m_\pm(n, c)$ such that*

$$\lceil m_-(n,c) \rceil \leqslant m^* \leqslant \lfloor m_+(n,c) \rfloor, \tag{9}$$

with

$$m_\pm(n,c) = \frac{1}{8c^4}\left((8n-9)c^4 - 2nc^3 + n(n-2)c^2 \pm 2c^2\sqrt{n(n-1)\left((8n-9)c^2 - n(2c+1)\right)}\right).$$

(iii) *If $c \in [c_1^*, c_2^*]$, then*

[15] $f(n) \sim g(n)$ as $n \to \infty$ if $\lim_{n\to\infty} \frac{f(n)}{g(n)} = 1$.
[16] Proposition 1 extends the basic efficiency results obtained in König et al. (2008), which uses a similar knowledge accumulation process as in this paper.
[17] It is possible to construct an interlinked star not having a nested neighborhood structure. As an example, consider the graph composed of three cliques of increasing size (from one to three) and a hub being connected to all nodes. Note that the node with minimal degree (the clique of size one) is connected to the node having maximal degree (the hub). This graph is an interlinked star. However, it easy to see that this graph does not have a nested neighborhood structure and thus it is not a nested split graph. By construction, the neighborhood of the nodes in the clique of size two includes the nodes in that clique. But these nodes are not in the neighborhood of the nodes in the clique of size three.
[18] $\lceil x \rceil$, denotes the smallest integer larger or equal than x (the ceiling of x, with $x \in \mathbb{R}$). Similarly, $\lfloor x \rfloor$ denotes the largest integer smaller or equal than x (the floor of x).

$$0 \leqslant m^* \leqslant \left\lfloor \frac{n(n-1)}{2c^2} \right\rfloor. \tag{10}$$

(iv) *If* $c_2^* < c$, *then* $m^* = 0$.

Both, the upper and lower bounds in items (ii) and (iii) are decreasing with cost c (see also the proof of Proposition 2 in Appendix A). This indicates that above $c = 0.5$ the efficient network tends to be more sparse as the cost of forming links increases. In particular, in item (ii) when n is taken to infinity, the number of links in the efficient network is approximately $\frac{n^2}{8c^2}$, which is close to the one of the complete network when $c = 0.5$.

A further specification of the efficient structure would require the determination of the graph that maximizes the largest eigenvalue λ_{PF} within the class of nested split graphs. At this point it is relevant to mention the relation with the efficient structure emerging in models in which agents maximize their centrality (Ballester et al., 2006). More precisely, Corbo et al. (2006) show that, in such a framework, the networks that maximize welfare are also the graphs with maximal eigenvalue. However, there, the number of nodes and edges of the efficient graph G^* is subject to constraints. If we imposed the same constraints as in Corbo et al. (2006), we would obtain the same (unique) efficient graphs. In contrast, in our model there is no restriction on the number of edges. In this case, finding the graph that maximizes the largest eigenvalue within the class of connected nested split graph is an unsolved problem in spectral graph theory (Stevanovic, 2007). So far, results are only available for specific values of the number of nodes (see also below).

Lemma 2 and Propositions 1 and 2 deliver important implications for the literature. First, they refine previous results on efficiency by Goyal and Joshi (2003) and Westbrock (2010). Indeed, both papers find that interlinked stars are efficient for positive costs of link formation. We find that a subset of interlinked stars, namely the asymmetric connected nested split graphs, can be efficient for $\frac{n}{2n-3} < c < \infty$ ($\sim [0.5, \infty)$ for large n) and are surely efficient if $c \in [\frac{n}{2n-3}, 1)$ ($\sim [0.5, 1)$ for large n). Furthermore, Westbrock (2010) shows that asymmetric efficient structures are efficient under direct spillovers and industry-wide indirect spillovers, and that they can either have a connected (interlinked-star) or disconnected (dominant-group) architecture. We extend this result to the case of indirect network-dependent spillovers. In our model efficient asymmetric networks are surely connected for $\frac{1}{2} < c < 1$, and in this case they are a subset of interlinked stars (connected nested split architecture). In addition, they can be disconnected for $1 < c < \sqrt{\frac{n(n-1)}{2}}$. However, in the latter case efficient structures have at most one non-trivial component. This is similar to the results obtained by Westbrock (2010), where the only disconnected structure is the dominant-group architecture.

The similarity between our results and those of Westbrock's analysis is explained by the degree variance of the efficient graphs in the two models. On the one hand, Westbrock (2010) finds that efficient networks are those maximizing the (normalized) degree variance (cf. Proposition 2 in Westbrock's paper). On the other hand, the networks maximizing degree variance (either connected or disconnected) belong to the class of nested split graphs (Hagberg, 2004). Since our efficient graphs also belong to this class,[19] it follows that they can display a structure similar to those in Westbrock's model. However, despite the aforementioned similarities, it must be stressed that the economic mechanism leading to the efficiency of asymmetric structures is very different in our model. In Westbrock (2010), the market setting plays a crucial role in the generation of efficient asymmetric structures. This is because, social returns from a collaboration are increasing in firm's scale and firm's scale is increasing in firm's degree. Thus, it is socially desirable to reallocate a link from a firm having few connections to one with many connections. In our model instead, firm's scale plays no role. Social returns to link formation depend on the ability of a collaboration to increase the number of walks in the network because walks drive indirect spillovers and ultimately innovation and profits in the model. Concentrating the number of collaborations among few firms is socially desirable inasmuch as it contributes to increase the number of walks while saving on the costs of direct collaborations. More precisely, in this model a social planner is confronted with the need of structuring the network in order to create the highest possible number of walks across firms while saving as much as possible on the number of direct connections. When the cost of collaboration is low, so is the cost of maintaining indirect walks through many direct connections. This explains why efficient networks are dense and symmetric for low values of collaborations costs. However, as the cost of collaboration increases, it is better for the industry to economize on the number of direct connections and to sustain indirect walks by concentrating direct connections among few firms. In this regard, two major transitions occur in our model as collaboration costs increase. Above 0.5 ($c > \frac{n}{2n-3}$ to be precise) the efficient graph is not complete anymore and becomes asymmetric. Above 1, it may break down and become disconnected. However, if it is disconnected, it always contains at most one non-trivial connected component, while the remaining nodes are isolated. This is because, in order to maintain a high number of walks while economizing on direct collaborations, it is efficient to completely exclude some firms from the network rather than breaking down the network into smaller components having non-trivial structures.

Finally, efficient structures in our model are more complex than others typically found in the network formation literature. For instance, differently from the model of Bala and Goyal (2000), the efficient graph is in general not minimally connected (removing one link does not necessarily make the graph disconnected). In addition, differently from Jackson and

[19] However, the graphs maximizing the eigenvalue and the graphs maximizing degree variance do not always coincide, as it has been shown by Bell (1992).

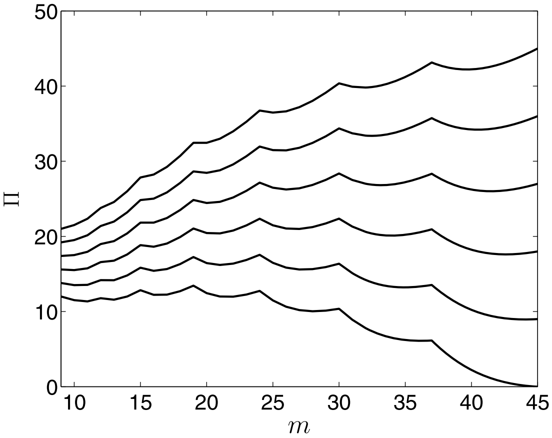


Fig. 2. Aggregate profits $\Pi(G)$ for the graphs which maximize the largest real eigenvalue λ_{PF} among all connected graphs with $n = 10$ nodes for all possible number of links m and cost $c \in \{0.5, 0.6, \ldots, 1\}$. The higher the cost c the lower the curve indicating aggregate profits $\Pi(G)$. The local maxima are always attained for the graph $F_{n,d}$.

Wolinsky (1996), in our model the set of efficient graphs is not limited to the star and the complete graph, but it includes a whole class of graphs that can be seen as intermediate graphs between these two extreme cases.

Not only the architecture but also the value of total profits is of interest if, for instance, one wants to design a network that performs close to efficiency. The value of total profits associated with a connected efficient graph G^* can be approximated by the total profits associated with the nested star $F_{n,d}$ (cf. Section 2.1), as stated in the next proposition.[20]

Proposition 3. *Let G^* be the efficient graph for a given number n of firms, and $c \geqslant 0$ the cost of collaboration. Denote the relative error of total profits between G^* and the nested star $F_{n,d}$ by $\epsilon = (\Pi(G^*, c) - \Pi(F_{n,d^*}, c))/\Pi(G^*, c)$, where $d^* = \arg\max_d \Pi(F_{n,d}, c)$. Then ϵ is bounded by*

$$\epsilon \leqslant \frac{c^2(8n-9) - n - 2nc}{n^2 + 2nc - c^2(8n-9)},$$

implying that $\epsilon \to 0$ in the limit of large n.

Notice that the above proposition focuses on the relative error since, by definition, total profits of both $F_{n,d}$ and G^* grow with the size n of the industry. Moreover, the relative error is small already for moderate values of n.[21]

Fig. 2 shows maximum aggregate profits for $n = 10$, any number of links $n - 1 \leqslant m \leqslant \binom{n}{2}$ and specific values of cost c. The figure shows that the local and global maxima of aggregate profits are always attained by the graph $F_{n,d}$.[22] It is worth noticing that it has been conjectured that the class of graphs that maximize λ_{PF}, is a subclass of the connected nested split graphs containing the nested stars $F_{n,d}$ (Aouchiche et al., 2008). However, this remains for the moment only a conjecture and all one can prove is that there exist specific combinations of n (e.g. $n = 10$, see Aouchiche et al., 2008) and c such that $F_{n,d}$ maximizes aggregate profits. Fig. 3 shows examples of nested stars $F_{n,d}$ for industry size $n = 10$ and values of cost c for which $F_{n,d}$ returns maximal aggregate profits. Consistently with Propositions 1 and 2 and previous discussions, the figure reveals that, as c increases, the network becomes more sparse, a core–periphery structure emerges and links are more concentrated (all peripheral nodes connect to a single hub in the core).

4. Stability

The analysis contained in the previous section studies which network topologies maximize industry welfare. In this section, we turn to the investigation of the existence and properties of equilibrium network structures. To this end, as in Goyal and Joshi (2003), Goyal and Moraga-Gonzalez (2001) and Westbrock (2010), we adopt the criterion of pairwise stability (cf. Jackson and Wolinsky, 1996).

[20] Since the lower bound on the eigenvalue of the nested star $F_{n,d}$ employed in the proof uses the largest eigenvalue of K_{d^*}, total profits of any architecture containing a clique of size d^* will tend, for large n, to those of the efficient graph. Indeed, for large n, the contribution to total profits provided by the $n - d^*$ links tends to 0. One could, for instance, connect $n - d^*$ nodes in a path instead of connecting them to a single node as in a star. However, among the possible alternative architectures, $F_{n,d}$ is the only graph that has a stepwise adjacency matrix. $F_{n,d}$ therefore has a higher eigenvalue and hence also higher aggregate profits.

[21] For example with $n = 100$ we get an error below 2%, while for $n = 200$ the error is below 1% with a marginal cost of collaboration $c \in [0, 1)$.

[22] Note that in the case of $F_{n,d}$, the number of links is given by $m = \binom{d}{2} + (n - d)$.

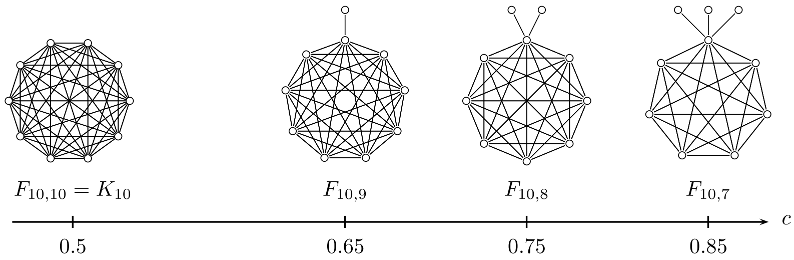


Fig. 3. Examples of graphs $F_{n,d}$ for values of cost $c = 0.5, 0.65, 0.75, 0.85$ and $n = 10$. *Note*: For these specific values of cost c and industry size n, $F_{n,d}$ is the graph that returns maximal total profits.

Definition 1. The graph G is *pairwise stable* if

(i) $\forall ij \in E(G)$, $\pi_i(G) \geqslant \pi_i(G - ij)$ and $\pi_j(G) \geqslant \pi_j(G - ij)$;
(ii) $\forall ij \notin E(G)$, if $\pi_i(G + ij) > \pi_i(G)$ then $\pi_j(G + ij) < \pi_j(G)$, and, if $\pi_j(G + ij) > \pi_j(G)$ then $\pi_i(G + ij) < \pi_i(G)$.

In light of the foregoing definition, we now proceed to investigate the stability of specific network structures. A straightforward application of Lemma 1 implies that, on the one hand, when the marginal cost c is larger than one, the marginal gross payoff is always lower than the marginal cost (cf. item (iii) of Lemma 1) and the only equilibrium is the empty graph $\bar{K}_n$.[23] On the other hand, when marginal costs are zero, links are always profitable, no existing link is deleted and thus the unique equilibrium is the complete graph K_n. Besides the foregoing extremes, the determination of stable networks becomes quite involved. This is because, in general, the marginal gross payoff from collaborations depends on the *topology* of the graph and on the *position* of the firm in the network (cf. Section 2.3). However, in line with previous work in the strategic network formation literature, it is possible to characterize the stability of various network structures in the space (c, n) of collaboration cost c and industry size n, as it is carried out in the following proposition[24]:

Proposition 4. *Let c denote the marginal cost of a collaboration and n the total number of firms in the R&D network G. Then the stability conditions for the different types of graphs identify the following regions in the parameter space $(c, n) \in \mathbb{R}_+ \times \mathbb{N}$:*

(i) *For cost $c > 1$ (and any n) the empty graph $\bar{K}_n$ is the unique stable network.*
(ii) *The complete graph K_n is stable if and only if*

$$n < \left\lfloor \frac{2 - c(1 - c)}{c} \right\rfloor. \tag{11}$$

If costs are zero, $c = 0$, then the complete graph K_n is the unique stable network (for any n).
(iii) *The graph consisting of a set of $d \geqslant 2$ equally sized, disconnected cliques $K_k^1, K_k^2, \ldots, K_k^d$ (G having $n = dk$ nodes in total) is stable if there exists an integer $k < n$, with $\mathrm{mod}(n, k) = 0$ such that*

$$\left\lceil \frac{1 + c(1 - c)}{c} \right\rceil \leqslant k \leqslant \left\lfloor \frac{2 - c(1 - c)}{c} \right\rfloor. \tag{12}$$

(iv) *The star $K_{1,n-1}$ is stable if*

$$\left\lceil \frac{2}{c} \right\rceil \leqslant n \leqslant \left\lfloor \frac{1 + c^2(6 + c^2)}{4c^2} \right\rfloor. \tag{13}$$

(v) *The graph consisting of $d \geqslant 2$ disconnected cliques $K_k^1, \ldots, K_k^d$ with $n = kd$, and the star $K_{1,n-1}$ are both stable for all pairs of integers k, n with $k \leqslant n$ such that the conditions in Eqs. (12) and (13) hold.*

[23] For $c = 1$ the empty graph is stable but not unique. Other structures can be stable (e.g. a set of dyads). Moreover, it should be noted that in some instances the decline in the eigenvalue can be even larger than one for some members of a component (for example of a star or of unequal connected cliques). This implies that the unilateral link deletion underlying the concept of pairwise stability is a necessary condition for the uniqueness of the empty and of the complete graph.

[24] In Appendix A we report only the proofs of items (vi) and (vii) of Proposition 4. For the proof of item (i) see the proof of Proposition 4 in König et al. (2011). Moreover, the stability of the complete graph K_n follows directly from the proof of Proposition 8 in König et al. (2011), while uniqueness has been shown in Proposition 3 in König et al. (2011). Furthermore, the proofs of item (iii) and (iv) are analogous to the proofs of the more general Propositions 5 and 6 in König et al. (2011). Finally, as far as item (v) is concerned, by means of numerical computations we could identify multiple pairs (c, n) for which both, the star and the disconnected cliques are stable. A version of Fig. 4 in which the pairs of (c, n) where stable cliques exist are indicated can be obtained upon request from the authors.

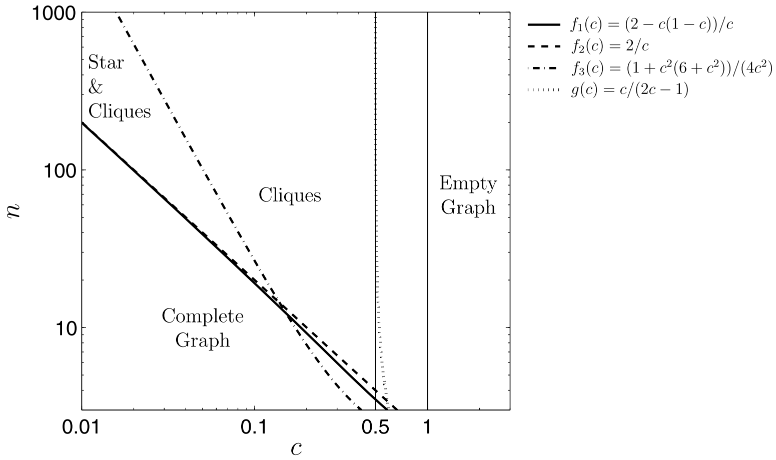


Fig. 4. Characterization of stable networks for combinations of cost c and network size n. In order to make visible the intersections among the curves, both axes are in logarithmic scale. The solid, dashed and dot-dashed curves correspond to the bounds in Eqs. (11) and (13), respectively. The dotted curve corresponds to the bound in Proposition 1, item (ii).

(vi) *There exists a range of cost* $c_1(n) < c < c_2(n) < 1$ *such that the dominant-group architecture* $D_{n,k}$ *is stable if* $n = k + 1$ *but it is not stable if* $n \geqslant k + 2$.

(vii) *Any network with two or more isolated nodes is not stable in the cost range* $c \in [0, 1)$.

Fig. 4 illustrates the results of Proposition 4 in the space of cost c and industry size n.[25] As the diagram reveals, in the region to the right of the vertical line at $c = 1$ ($c > 1$ and any network size n), the empty graph $\bar{K}_n$ is the unique stable network (*empty graph region*). Below the solid curve $f_1(c)$, corresponding to inequality (11), the complete graph K_n is stable (*complete graph region*), while above this curve it is not.[26] In the region above the same solid curve there exist combinations of (c, n) for which graphs comprised of multiple cliques are stable (*multiple-cliques region*), corresponding to divisors k of n satisfying the conditions (12). Clearly, the larger is n, the more numerous are the suitable divisors. Thus, in this region the number of equilibria grows with n.[27] Notice that the upper bound for the cliques size in (12) decreases with cost c (for $0 < c < 2$), while the lower bound increases for decreasing cost. This implies that clique size decreases with cost c, as confirmed also by numerical computations (see footnote 27). For instance, for $c > 0.618$, we find that stable cliques contain no more than 2 nodes (dyads). Besides, a triangular area is delimited above the dashed curve $f_2(c)$ and below the dashed–dotted curve $f_3(c)$ according to the inequalities in (13). In this region, the star $K_{1,n-1}$ is stable (*star region*). Note that this region partially overlaps with the region in which multiple cliques are stable. It follows that there exist combinations of cost c and size n for which both the multiple cliques and the star can be stable. Finally, notice that in the diagram there are no boundaries concerning the dominant architecture because this architecture is not stable in $c \in [0, 1)$ except for the special case $n = k + 1$.

To summarize, for any $c \in [0, 1)$ the following applies. For small n, the complete graph K_n is a stable network, while for large n, the complete graph is not stable and other structures such as the star and the multiple cliques can be stable.[28] Moreover, there is a region, corresponding to small values of c and intermediate values of n, in which both the star and the multiple cliques can be stable. The intuition for these results exploits the property (see also Section 2.3) that in all the architectures of Proposition 4, the marginal revenue from collaborations (i.e. $\Delta\lambda_{\mathrm{PF}}$) decreases with the size of the connected component to which the firm belongs. Since the eigenvalue is related to the number of walks (cf. Section 2.3), this property can be interpreted as the fact that the contribution of one link to the overall number of walks in a connected component decreases with its size. A simple formal argument for this intuition can be provided for the following case. Consider a clique of size n and the formation of n links to a previously disconnected node, in order to obtain a new clique of size $n + 1$.

---

[25] In König et al. (2011) we prove that the stability of the cliques and of the star is robust to the introduction of a cost asymmetry between link creation and link deletion.

[26] Notice that the complete graph is *efficient* in the whole region to the left of the dotted curve $g(c)$ corresponding to the bound in Proposition 1, item (ii). See also Proposition 5.

[27] The ceiling and floor functions limit the values of c where such k may exist. For example, when $c = 0.6$ no k satisfies the conditions. Notice that this is related to the discrete nature of k rather than to a substantial incentive problem. Also, once a value of k satisfying Eq. (12) exists, then there is an infinite number of stable networks with varying n (the number of cliques can be multiplied indefinitely). In order to find out how frequent these configurations are, we carried out numerical computations for all integers $n \in \{1, 2, \ldots, 1000\}$ and for values of cost $c \in \{0, 0.01, 0.02, \ldots, 1\}$. For each resulting combination (c, n) we examined whether suitable divisors of n exist. For almost all values of c, we find that cliques are stable for a large fraction of values of n. For the sake of readability these results are not shown in the diagram. They are available from the authors upon request.

[28] The lower bound in item (iv) of Proposition 4 is not tight. Thus we cannot rule out the existence of stable stars also in the region wherein the complete graph is stable. See also the proof of the proposition.

The total change in the largest eigenvalue is $\Delta\lambda_{\mathrm{PF}} = 1$ (see also Table 1), while the change in number of links is $\Delta m = n$. Therefore, the average contribution to the increase in eigenvalue provided by each link is $\frac{\Delta\lambda_{\mathrm{PF}}}{\Delta m} = \frac{1}{n}$. It follows that, for any given c and for n large enough, $\frac{\Delta\lambda_{\mathrm{PF}}}{\Delta m} < c$. This means for large enough networks, at least one node in the clique does not have the incentive to create a link. The change in eigenvalue is the same in the case of an exclusion of one node from a clique of size $n+1$, which implies that firms have a stronger incentive to severe a link in large enough networks.[29]

The above intuition explains why the complete network is an equilibrium only in small industries (and for small costs c). It explains also why, in industries of growing size, walks are sustained, instead, by forming either asymmetric connected structures (the star) or disconnected structures (the cliques), characterized by high density within-components. The incentive to form connections in industries of growing size further decreases with marginal cost c. That is why, when both marginal cost and industry size increase, we observe that only disconnected cliques of decreasing size are stable. Marginal revenues decreasing with component size also explain why all networks with more than two isolated nodes (and the dominant-group architecture in particular) are not stable in this model. Indeed, the change in the eigenvalue from establishing a connection between two isolated firms is the highest possible ($\lambda_{\mathrm{PF}} = 1$, see also proof of Proposition 4), and for $c < 1$ these firms always have the incentive to form a collaboration.

Proposition 4 and the stability diagram of Fig. 4 deliver several implications in relation to the literature on R&D networks. For instance, Goyal and Joshi (2003) and Westbrock (2010) show the stability of asymmetric network structures (the dominant-group architecture) in the presence of direct spillovers and industry-wide indirect spillovers. Moreover, network asymmetry is often observed in practice (see e.g. Powell et al., 2005). On one hand, Proposition 4 confirms that asymmetric networks can be stable also in presence of network-dependent indirect spillovers. In particular, asymmetric equilibrium networks emerge as a characteristic of industries with large size (i.e. n high) and wherein costs of collaboration are small. On the other hand, asymmetric equilibrium structures are not a general feature with network-dependent indirect spillovers. First, in the same region in which asymmetric structures are stable,[30] there exist combinations of industry size n and cost c for which perfectly symmetric and disconnected structures (the cliques) can be stable. Second, symmetric (and increasingly sparser) structures are stable as the cost of collaboration increases while the star is not stable anymore.

Finally, stable networks in our model (both symmetric and asymmetric) do not feature, in general, firms that are excluded from any collaborative activity. Indeed, the dominant-group architecture is never stable in our model except for the special case of a clique plus one isolated node, which is stable for some combinations of n and $c \in [0, 1)$.

Our results have also implications for the game theoretic literature on network formation. First, stable graphs are not always minimally connected (e.g., in the multiple clique equilibrium each component is complete). This is an important feature that, for instance, distinguishes our results from the linear "two-way flow" model of Bala and Goyal (2000). Furthermore, the fact that disconnected equally sized cliques are stable is a specific property of our model,[31] since disconnected networks are not generally pairwise stable in network formation models with positive linking externalities. They are instead common in models with negative linking externalities (e.g. the "co-author" model in Jackson and Wolinsky, 1996). Finally, our results also contrast with the predictions of local spillovers games studied in Goyal and Joshi (2006). In these models the existence of symmetric and asymmetric networks is found assuming strong monotonicity of agents' payoff in the creation/deletion of a link and under various hypotheses on the type of neighbors' spillovers and on the convexity/concavity of the agent's payoff function. In addition, symmetric and asymmetric networks are never stable for the same level of the model's parameters. In our model, symmetric and asymmetric stable networks emerge under very mild conditions on agents' payoffs (see Lemma 1), and can be stable for the same values of the model's parameters.

5. Stability vs. efficiency

In the previous section we have analyzed the stability of different network structures depending on industry size n and marginal costs of collaboration c. We now turn to the comparison of stability with efficiency. In this regard, the analysis of the stability diagram in Fig. 4 reveals a region of small industry size and small cost of collaboration in which stable graphs are also efficient. Indeed, to the left of the curve $g(c)$, the complete graph is the unique efficient graph (Proposition 1, item (ii)). At the same time, below the solid curve $f_1(c)$, the complete graph is also stable (Proposition 4, item (ii)). In

[29] A similar relation holds also for the other networks analyzed in Proposition 4. However, the formal argument is in this case more complicated, due to the characteristics of the relation between the eigenvalue and component size, and we therefore refer the reader to the proof of the proposition. Moreover, there is no general closed form expression for $\Delta\lambda_{\mathrm{PF}}$ as a function of the component size, because the change in eigenvalue typically depends on the specific network structure and not only on its size. Finally, the negative correlation between marginal revenues from collaborations and size must be distinguished from "concavity in own links" (e.g. Calvó-Armengol and Ilkilic, 2009; Hellmann and Buechel, 2009). This distinction is important because Hellmann and Buechel (2009) show that concavity in own links may generate under-connected, inefficient networks in models with positive linking externalities (as ours). Concavity in own links implies diminishing marginal revenues of a link with the degree of the agent. The relation stated above is instead between the marginal revenue and the number of agents in a component. In addition, note that the degree of some nodes grows with component size in some networks structures (this holds e.g., for each node in the clique, and for the hub in the star). However, in our model marginal revenues can decrease with component size even if the degree of a node does not change (e.g. for peripheral nodes in a star as additional nodes are attached to the hub; see the proof of Proposition 6 in König et al., 2011).

[30] See also the "connections model" by Jackson and Wolinsky (1996) for a similar result on the co-existence of symmetric and asymmetric stable networks under the same parameter values.

[31] See also Deroian (2008) for similar results on the stability of cliques without linking costs.

Table 1
Efficiency and stability of different types of networks.

Graph class	Eigenvalue	Efficiency	Stability
Empty graph $\bar{K}_n$	$\lambda_{\mathrm{PF}}=0$	$c>\sqrt{\frac{n(n-1)}{2}}$	$c>1$
Complete graph K_n	$\lambda_{\mathrm{PF}}=n-1$	$c\leqslant\frac{1}{2}$	$n\leqslant\lfloor\frac{2-c(1-c)}{c}\rfloor$
k-Symmetric graph	$\lambda_{\mathrm{PF}}=k-1$	if $k=n$, see K_n	see cliques or K_n
Star $K_{1,n-1}$	$\lambda_{\mathrm{PF}}=\sqrt{n-1}$	see CNS	$\lceil\frac{2}{c}\rceil\leqslant n\leqslant\lfloor\frac{1+c^2(6+c^2)}{4c^2}\rfloor$
Nested star $F_{n,d}$	$\lambda_{\mathrm{PF}}\geqslant d-1$	with good approx.	not for n large
Conn. nested split CNS	$\lambda_{\mathrm{PF}}\geqslant\sqrt{n-1}$	$0<c<\sqrt{\frac{n(n-1)}{2}}$	not for n large
Dominant-group $D_{n,k}$	$\lambda_{\mathrm{PF}}=\sqrt{k-1}$	if $k=n$, see K_n	not for $0\leqslant c<1$ if $n\geqslant k+2$
Cliques $\{K_d^1,\dots,K_d^l\}$	$\lambda_{\mathrm{PF}}=d-1$	if $l=1, d=n$, see K_n	$\lceil\frac{1+c(1-c)}{c}\rceil\leqslant d\leqslant\lfloor\frac{2-c(1-c)}{c}\rfloor$

contrast, the efficient graph does not belong to the set of possible equilibria if the size of the industry is large enough, as shown in the following proposition.

Proposition 5. *The efficient network G^* is not stable*

(i) *if $c\leqslant 0.5$ and $n>\frac{2-c(1-c)}{c}$, or if $c>0.5$ and $\frac{2-c(1-c)}{c}<n\leqslant\frac{c}{2c-1}$, or*
(ii) *if $c>0.5$ and $n\to\infty$, or*
(iii) *if $1<c<\frac{n+2\sqrt{2n^2(n-1)}}{8n-9}$ (with $n>3$).*

Item (i) follows directly from Propositions 1 and 4, and concerns the complete graph in the region of the stability diagram above curve $f_1(c)$ and to the left of curve $g(c)$ in Fig. 4. Notice that for $c>0.603$, the complete network is never stable as there is no n that satisfies the second condition in (i). Moreover, item (ii) shows the instability of a generic connected nested split graph, which is efficient for $c\in[0,1)$ – see Proposition 1, item (i) – when $c>0.5$ and $n\to\infty$. Finally, item (iii) concerns the region where the empty graph is the unique stable graph ($c>1$).[32]

Overall, the above proposition shows that the efficient graph is never stable in large industries ($n\to\infty$), for any given $0<c<\infty$, except for some extreme cases.[33] The instability of the complete and the connected nested split graph follows from the fact that the marginal revenue from collaborations (i.e. $\Delta\lambda_{\mathrm{PF}}$) decreases with component size, while the marginal cost of collaborations stays constant with size (see also the previous section). The number of links needed to maintain connected efficient network structures increases in large industries (for a network to remain connected, the number of links must be at least $n-1$). On the other hand, decreasing marginal revenues imply that in large industries firms have stronger incentives to severe collaborations and to make such efficient networks unstable.

Thus, in line with previous findings (e.g. Bala and Goyal, 2000; Jackson and Wolinsky, 1996; Westbrock, 2010), in our model the tension between efficiency and stability can be very pronounced. In particular, Westbrock (2010) focuses on network asymmetry, and finds that in large industries stable R&D networks are typically too little asymmetric with respect to the level that is socially desirable. Our results indicate that in presence of network-dependent indirect spillovers efficiency is also not attained in large industries ($n\to\infty$). Moreover, they indicate that connectedness is another important property in relation to efficiency. In particular, the fact that firms have stronger incentives to severe collaborations in large industries indicates that in these industries equilibrium graphs will be "under-connected" with respect to the level that is socially desirable.

Table 1 summarizes the results on efficiency and stability discussed so far and compares them with results for other well-known classes of graphs in the literature. The empty graph $\bar{K}_n$, the complete graph K_n, and the connected nested split graph (CNS) have already been discussed in Proposition 5. Two other graphs that deserve special attention are the star $K_{1,n-1}$, and the nested star $F_{n,d}$. The star is stable for values of c and n for which the complete graph is efficient. In addition, it can be efficient as a special case of asymmetric connected nested split graphs when $c>0.5$ (although, in this case it is not stable for $n\to\infty$, cf. Proposition 5).

The nested star $F_{n,d}$ is, too, a particular instance of a connected nested split graph. For this graph, it is possible to obtain a more restrictive condition on instability, because for $n\to\infty$ this graph is unstable for all values of cost $c>0$ (as opposed to $c>0.5$).[34] In the case of finite n, numerical computations suggest that for cost $c>\sqrt{2}-1=0.414$, the graph is unstable

[32] Indeed, for some values of costs $c>1$ the efficient graph is non-empty because the lower bound on the number of links in the efficient graph is positive (cf. Proposition 2, item (iii)). This implies that there exist values of costs $c>1$ for which the empty network is not efficient. In addition, as the proposition also shows, as n increases, values of $c>1$ for which the empty graph is not efficient also increase. However, for $c\to\infty$ and any n the empty graph is both efficient and stable.

[33] Indeed, notice that for any given $n<\infty$ the graph is efficient and stable for $c\to 0$. In addition, for any given $n<\infty$, the graph is always efficient and stable if $c\to\infty$.

[34] Bell (1991) provides an expression for $\lambda_{\mathrm{PF}}(F_{n,d})$ in terms of leading orders in n given by $\lambda_{\mathrm{PF}}(F_{n,d})=\sqrt{n}-\frac{1}{2\sqrt{n}}+\binom{d-1}{2}\frac{1}{n}+\frac{1}{8}(4d^3-20d^2+32d-17)\frac{1}{n^{\frac{3}{2}}}+O(\frac{1}{n^2})$. From this expression it is clear that for a given value of d, $\lim_{n\to\infty}(\lambda_{\mathrm{PF}}(F_{n,d})-\lambda_{\mathrm{PF}}(F_{n-1,d}))=0$ (corresponding to the change in eigenvalue

for any size $n \geqslant 3$.[35] For values of cost $c < 0.414$ and for n small we cannot exclude the existence of some stable $F_{n,d}$. However, for instance for $d \geqslant 10$ we find that $\Delta\lambda_{\mathrm{PF}}(F_{n,d}) < c$ for all values of $n \geqslant d$ whenever $c > 0.035$. This implies that, even for a small network comprising a clique of 10 nodes plus some peripheral nodes, the nested star $F_{n,d}$ is unstable unless the cost is very small (i.e. below 0.035).

Next, we consider the k-symmetric graph, i.e. the graph in which all nodes have the same degree k. For a given n, aggregate profits of this graph increase (decrease) with k if c is lower (higher) than one.[36] They are positive if $c < 1$, $k > \frac{1}{1-c}$ and this graph is efficient if $k = n-1$ and $c \leqslant 0.5$. This is in line with our results on efficiency (cf. Section 3), since the complete graph K_n is a particular case of a k-symmetric graph (when $k = n-1$). Note that the set of cliques of identical size k is also a particular case of $(k-1)$-symmetric graph. Finally, the dominant-group architecture is never stable in our model, except in the case of a clique and an isolated node. In addition, it is never efficient for $c \in [0, 1)$.

6. Conclusions

In this paper, we developed a model in which firms earn profits by recombining their knowledge in a network of R&D collaborations. We showed that under mild conditions on the rate at which firms discount profits from R&D collaborations, the gross payoff from collaborations depends on indirect network spillovers and is proportional to the growth rate of all walks existing across firms in their connected component. In addition, both individual and industry net payoffs display a trade-off between the benefits of the network walks generated by direct connections and the costs of direct connections.

In this framework, we characterized the topology of the efficient graph in relation to the marginal cost of collaboration. We showed that when the marginal cost of maintaining collaborations is low, the trade-off between walks and cost of direct connections is weak and the efficient network is the complete graph. For intermediate cost of collaborations the trade-off becomes tighter and the efficient graph belongs to the class of asymmetric connected nested split graphs, having a nested neighborhood structure. As costs get larger, the efficient graph can become disconnected and then empty.

Furthermore, we studied the existence of equilibrium graphs in the model, and the relation between equilibrium and efficiency. In particular, we showed that the complete graph is stable in small industries and for low collaboration costs, while the class of size-homogeneous disconnected cliques and the star are stable in large industries. We also showed that these two equilibrium structures may be stable for the same level of cost and that the size of stable cliques is decreasing in the cost of link formation. Finally, large industries, in our model, are characterized by a strong tension between efficiency and stability. This is because, the marginal benefits from R&D collaborations decrease with the size of the connected component of the firm, whereas marginal cost of collaboration remains constant with size. This implies that firms have smaller incentives to form collaborations in large industries, and therefore that equilibrium networks in those industries will be less connected with respect to what is efficient.

The present work could be extended in at least four directions. First, the model could be extended to account for industry demand and competition across firms, for example as in Goyal and Moraga-Gonzalez (2001). Second, gross benefits of firms from forming links in our model are determined by the largest real eigenvalue of their connected component. The largest eigenvalue is related to the growth of walks in the component but does not take into account the length of these walks. Hence, one could extend our model by introducing a utility function that discounts walks by their length (as it is the case for eigenvector or Bonacich centrality, see Ballester et al., 2006). This amounts to considering higher order terms in the payoff function of firms. Third, one could extend the analysis of network formation and study the different basins of attractions of the stable structures in our model, for example employing the simulation techniques suggested in Vega-Redondo (2007). Finally, one could investigate the robustness of equilibrium multiplicity in our model (and the relation between efficiency and equilibrium) using definitions alternative to pairwise stability e.g. of the kind proposed in Bloch and Jackson (2006).

Acknowledgments

We would like to thank Francis Bloch, Matthew Jackson, Giorgio Fagiolo, Hans Haller, Luigi Marengo, Koen Frenken, Paolo Saviotti, Nicolas Carayol, Patrick Groeber, Hans Gersbach, Douglas White and two anonymous referees for their comments. Moreover we would like to thank participants at the 12th International Conference of the Schumpeter Society in Rio de Janeiro, 2008, participants at the 14th International Conference on Computing in Economics and Finance in Paris, 2008 and the audience at the North American Winter Meeting of the Econometric Society in San Francisco, 2009, where earlier versions of this paper were presented. M.D. König acknowledges financial support from Swiss National Science Foundation through research grant PBEZP1-131169. All usual disclaimers apply.

Appendix A

In this appendix we give the proofs of the propositions and lemmas stated in the paper.

---

from disconnecting a peripheral node from the central node in $F_{n,d}$). Since the change in the largest eigenvalue tends to zero as n increases, $F_{n,d}$ becomes unstable for arbitrarily small values of cost $c > 0$.

[35] The results of these numerical computations are not reported here because of space constraints, but they are available upon request from the authors.

[36] Total profits of the k-symmetric graph are given by $\Pi = n\lambda_{\mathrm{PF}} - 2mc = n(k-1) - nkc = n(k(1-c)-1)$.

Proof of Lemma 2. (i) Since G_1 and G_2 are connected, we have that $m_1 \geqslant n_1 - 1$ and $m_2 \geqslant n_2 - 1$. We now consider all possible cases for the number of links in the two components and the resulting increase in aggregate profits from connecting them.

(a) First, we consider the case of $m_1 \geqslant n_1$ and $m_2 \geqslant n_2$. Assume that the largest eigenvalue of G_1 is $\lambda_{\mathrm{PF}}(G_1) \geqslant \lambda_{\mathrm{PF}}(G_2)$. Let G' be the graph obtained as follows. For each node in G_2 we rewire one incident link to a node in G_1. This means that each time we add a node from G_2 to the graph G_1 and a link connecting the added node to the rest of the connected component. In this way, all nodes in G_2 get connected to G_1. The number of rewired links is n_2 (and there are at least that many links, since by assumption $m_2 \geqslant n_2$). Adding a link to a connected graph strictly increases its eigenvalue, since the underlying adjacency matrix is irreducible (Horn and Johnson, 1990, p. 515). Therefore, $\lambda_{\mathrm{PF}}(G') > \lambda_{\mathrm{PF}}(G_1) \geqslant \lambda_{\mathrm{PF}}(G_2)$ and total profits of G' are

$$\Pi\big(G'\big) = (n_1+n_2)\lambda_{\mathrm{PF}}\big(G'\big) - 2(m_1+m_2)c > n_1\lambda_{\mathrm{PF}}(G_1) + n_2\lambda_{\mathrm{PF}}(G_2) - 2(m_1+m_2)c = \Pi(G_1) + \Pi(G_2). \tag{14}$$

(b) Let $m_1 \geqslant n_1$ and $m_2 = n_2 - 1$. Since $m_2 = n_2 - 1$ the largest real eigenvalue of G_2 is at most the one of the star K_{1,n_2-1}, hence $\lambda_{\mathrm{PF}}(G_2) \leqslant \lambda_{\mathrm{PF}}(K_{1,n_2-1}) = \sqrt{n_2-1}$ (Cvetkovic and Rowlinson, 1990). We first construct the graph G_1' from G_1 with the same number of nodes n_1 and links m_1 but whose adjacency matrix is stepwise. This can be achieved by moving the one entries to the left in the adjacency matrix of G_1. The resulting graph G_1' has the property that $\lambda_{\mathrm{PF}}(G_1') \geqslant \lambda_{\mathrm{PF}}(G_1)$ (Cvetkovic and Rowlinson, 1990). It holds that G_1' has a spanning star as a subgraph. Moreover, denote by G' the graph obtained by connecting all nodes in G_2 to the central node in the star in G_1'. For this we need to add one link to connect all n_2 nodes so that G' has n_1+n_2 nodes and $m_1+m_2+1 = m_1+n_2$ links. The resulting graph G' has again a stepwise matrix and contains a spanning subgraph K_{1,n_1+n_2-1}. Therefore, it must hold that $\lambda_{\mathrm{PF}}(G') > \lambda_{\mathrm{PF}}(K_{1,n_1+n_2-1}) = \sqrt{n_1+n_1-1}$. The inequality is strict because there is at least one link more than in K_{1,n_1+n_2-1} following from $m_1 > n_1 - 1$. It follows that $\Pi(G') = (n_1+n_2)\lambda_{\mathrm{PF}}(G') - 2(m_1+n_2)c$. Thus, we get

$$\begin{aligned}
\Pi\big(G'\big) - \big(\Pi(G_1) + \Pi(G_2)\big) &\geqslant \Pi\big(G'\big) - \big(\Pi(G_1) + \Pi(K_{1,n_2-1})\big) \\
&= n_1\big(\lambda_{\mathrm{PF}}\big(G'\big) - \lambda_{\mathrm{PF}}(G_1)\big) + n_2\big(\lambda_{\mathrm{PF}}\big(G'\big) - \sqrt{n_2-1}\big) - 2c \\
&> n_2\big(\lambda_{\mathrm{PF}}\big(G'\big) - \sqrt{n_2-1}\big) - 2 \\
&> n_2(\sqrt{n_1+n_2-1} - \sqrt{n_2-1}) - 2 \\
&\geqslant n_2(\sqrt{n_2+2} - \sqrt{n_2-1}) - 2,
\end{aligned} \tag{15}$$

where we have used the fact that $n_1 \geqslant 3$ in the last inequality from above (for $n_1 = 2$ or $n_1 = 1$ we cannot have $m_1 \geqslant n_1$ without introducing loops or multiple links, which we have ruled out by assumption). This inequality is non-negative for any value of $n_2 > 1$. Hence, we have shown that $\Pi(G') > \Pi(G_1) + \Pi(G_2)$. A similar result holds for $m_1 = n_1 - 1$ and $m_2 \geqslant n_2$.

(c) Assume that $m_1 = n_1 - 1$ and $m_2 = n_2 - 1$. In this case, both components must be stars, K_{1,n_1-1} and K_{1,n_2-1} with eigenvalues $\sqrt{n_1-1}$ and $\sqrt{n_2-1}$. Construct the graph G' by connecting n_2 nodes from K_{1,n_2-1} to the central node in K_{1,n_1-1}. The resulting graph is K_{1,n_1+n_2-1} and (assuming w.l.o.g. $n_1 \geqslant n_2$) it holds that

$$\begin{aligned}
\Pi\big(G'\big) - \big(\Pi(K_{1,n_2-1}) + \Pi(K_{1,n_2-1})\big) &= n_1(\sqrt{n_1+n_2-1} - \sqrt{n_1-1}) + n_2(\sqrt{n_1+n_2-1} - \sqrt{n_2-1}) - 2c \\
&> (n_1+n_2)(\sqrt{n_1+n_2-1} - \sqrt{n_1-1}) - 2,
\end{aligned} \tag{16}$$

for $c \in [0,1)$. The right-hand side is non-negative as long as $n_2 \geqslant 2$.

(d) Next, let $n_1 \geqslant 2$ and $n_2 = 1$. This case corresponds to a connected graph G_1 and one isolated node. Total profits are $\Pi(G) = n_1\lambda_{\mathrm{PF}}(G_1) - 2m_1c$. Denoting the graph G' obtained by connecting the isolated node to G_1, we find $\Pi(G') = (n_1+1)\lambda_{\mathrm{PF}}(G') - 2(m_1+1)c \geqslant \Pi(G) + (\lambda_{\mathrm{PF}}(G') - 2c)$. In order to show this we have to consider the cases $n_1 = 2, 3$ and $n_1 \geqslant 4$ separately. (1) If $n_1 \geqslant 4$, then $m_1 \geqslant n_1 - 1$ (since G_1 is connected by assumption). We can construct a star K_{1,n_1-1} plus additional edges from G_1 and connect the isolated node to it. Denote the resulting graph G'. Then, $\lambda_{\mathrm{PF}}(G') \geqslant \lambda_{\mathrm{PF}}(K_{1,n_1}) = \sqrt{n_1} \geqslant 2$ and it follows that $\Pi(G') - \Pi(G) \geqslant 0$ if $\lambda_{\mathrm{PF}}(G') \geqslant 2 > 2c$ for $c \in [0,1)$. (2) If $n_1 = 3$, then G_1 is either a path P_3 (with length 2) or a cycle C_3 containing 3 nodes. We connect the isolated node to G_1. In the case of $G_1 = P_3$ we get $\Pi(G') - \Pi(G) = (4\sqrt{3} - 6c) - (3\sqrt{2} - 4c) = 2.69 - 2c > 0$, where the last inequality follows from $c \in [0,1)$. In the case of $G_1 = C_3$ we obtain $\Pi(G') - \Pi(G) = (42.17 - 8c) - (32 - 6c) = 2.68 - 2c > 0$ again, using $c \in [0,1)$. (3) We consider the case $n_1 = 2$. We connect the isolated node to $G_1 = P_2$ and again denote the resulting connected graph G'. We then have that $\Pi(G') - \Pi(G) = (3\sqrt{2} - 4c) - (2 - 2c) = 2.24 - 2c > 0$, with $c \in [0,1)$.

(e) Finally, for the case $n_1 = 1$ and $n_2 = 1$, we have two isolated nodes with total profits $\Pi(G) = 0$. If we connect the nodes via an edge we have $\Pi(G') = 2(1-c)$. Since $c \in [0,1)$ total profits in the connected graph G' are higher. The above cases consider all possible disconnected graphs and show that total profits Π can be increased by connecting them.

(ii) The proof proceeds in a similar way as the one of part (i). Since G_1 and G_2 are connected, we have that $m_1 \geqslant n_1 - 1$ and $m_2 \geqslant n_2 - 1$. We now consider all possible cases for the number of links in the two components and the resulting increase in aggregate profits from connecting them.

The case of (a) $m_1 \geqslant n_1$, $m_2 \geqslant n_2$ has been treated already in the proof of Lemma 2, item (i). In the case of (b) $m_1 \geqslant n_1$, $m_2 = n_2 - 1$, we must have that $G_2 = K_{1,n_2-1}$. We first rewire the links in G_1 in the same way as in (i) item (b), in order to obtain a graph G_1' with the same number of nodes n_1 and links m_1 as in G_1, such that the adjacency matrix of G_1' is stepwise and has the largest eigenvalue among all graphs with that number of nodes and links. We then construct the graph G' by rewiring n_2-1 links in K_{1,n_2-1} to the node in G_1' with the highest degree. This makes G' a connected graph and one isolated node, and thus, falling into the class of graphs considered in the lemma. Observe also that G' must have a spanning star K_{1,n_1+n_2-2} as an induced subgraph, so that $\lambda_{\mathrm{PF}}(G') > \lambda_{\mathrm{PF}}(K_{1,n_1+n_2-2}) = \sqrt{n_1+n_2-2}$. Since nodes and links have been added to G_1', it must have a strictly larger eigenvalue than G_1. We then have that

$$\begin{aligned}\Pi\big(G'\big) - \big(\Pi(G_1)+\Pi(G_2)\big) &> n_1\lambda_{\mathrm{PF}}(G_1) + (n_2-1)\sqrt{n_1+n_2-2} - n_1\lambda_{\mathrm{PF}}(G_1) - n_2\sqrt{n_2-1} \\ &= (n_2-1)\sqrt{n_1+n_2-2} - n_2\sqrt{n_2-1}.\end{aligned}$$

The expression $(n_2-1)\sqrt{n_1+n_2-2} - n_2\sqrt{n_2-1}$ is strictly positive for any $n_1 > 3$ and $n_2 > 1$, and so G' has higher aggregate profit for any c. For $n_1 = 3$, the only possible graph satisfying $m_1 \geqslant n_1$ is the cycle C_3. In this case we can construct the graph K_{1,n_1+n_2-1}, which has a higher eigenvalue than both K_{1,n_2-1} and C_3. This graph is connected, and falls into the class of graphs considered in the lemma. When considering the case of (c) $m_1 = n_1-1 \geqslant 2$ and $m_2 = n_2-1 \geqslant 2$, where both components must be stars, K_{1,n_1-1} and K_{1,n_2-1}, with eigenvalues $\sqrt{n_1-1}$ and $\sqrt{n_2-1}$, the same proof can be used to show that $\Pi(G') > \Pi(G)$.

The above cases consider all possible disconnected graphs and show that total profits Π can be increased by connecting them while possibly leaving some nodes isolated. □

Proof of Proposition 1. (i) For the complete graph K_n we have that $\lambda_{\mathrm{PF}} = n-1$, $m = \frac{n(n-1)}{2}$ and total profits are $\Pi(K_n) = n(n-1) - 2\frac{n(n-1)}{2}c = n(n-1)(1-c)$. On the other hand, the largest real eigenvalue λ_{PF} of a graph G with m edges is bounded from above as $\lambda_{\mathrm{PF}} \leqslant \frac{1}{2}(\sqrt{8m+1}-1)$ (Cvetkovic and Rowlinson, 1990). For total profits we then have

$$\Pi = \sum_{i=1}^{n}\lambda_{\mathrm{PF}}(G_i) - 2mc \leqslant n\lambda_{\mathrm{PF}} - 2mc \leqslant \frac{n}{2}(\sqrt{8m+1}-1) - 2cm = b(n,m,c), \tag{17}$$

with $n \leqslant m \leqslant \binom{n}{2}$. For fixed cost c and number of nodes n, the number of edges maximizing $b(n,m,c)$, defined in Eq. (17), is given by $m^* = \frac{n^2-c^2}{8c^2}$ if $\frac{n^2-c^2}{8c^2} < \binom{n}{2}$ and $m^* = \frac{n(n-1)}{2}$ if $\frac{n^2-c^2}{8c^2} > \binom{n}{2}$. Note that the graph with the latter number of edges is the complete graph K_n. Inserting m^* into $b(n,m,c)$ yields

$$b\big(n,m^*,c\big) = \begin{cases} \frac{n}{2}\Big(\sqrt{\frac{n^2-c^2}{c^2}+1}-1\Big) - \frac{n^2-c^2}{4c}, & \text{if } c > \frac{n}{2n-1}, \\ n(n-1)(1-c) = \Pi(K_n), & \text{if } c < \frac{n}{2n-1}. \end{cases} \tag{18}$$

The bound for $c \leqslant \frac{n}{2n-1} \sim \frac{1}{2}$ in the limit of large n coincides with total profits of the complete graph K_n. Therefore K_n is the efficient graph in that region of cost. Also, all graphs with less edges than K_n have lower aggregate profits than K_n (the upper bound is decreasing with decreasing m) and so K_n is also the unique efficient graph in that region of cost.

(ii) We first prove that the efficient graph is connected when $c \in [0,1)$. For a contradiction assume that, for a given n, the efficient graph G is disconnected (and all connected graphs have smaller total profits than G). Then G must have at least two components. By part (i) of Lemma 2 each pair of components can be connected in such a way to result in a graph with higher total profits. Ultimately all components of G can be connected, yielding a connected graph G' with at least the total profits of G. This contradicts the assumption and it follows that the efficient graph must be connected. Further, among the connected graphs, the graphs with maximal eigenvalue have a stepwise adjacency matrix (Brualdi and Solheid, 1986). These graphs are referred to as connected nested split graphs (Aouchiche et al., 2008). Next, we show that the efficient graph is not complete. In Eq. (21) we derive a lower bound on aggregate profits $\Pi(F_{n,d},c)$ of the graph $F_{n,d}$. From this bound one can show that $\Pi(F_{n,d},c) \geqslant \Pi(K_n,c)$ if $c \geqslant \frac{n}{2n-3}$. This implies that the efficient graph G^* is not complete for $c \geqslant \frac{n}{2n-3}$. Moreover, from Lemma 2 we know that the efficient graph G^* is a nested split graph with possibly isolated nodes. This implies that G^* is not symmetric, since the complete graph is the only symmetric connected nested split graph.

(iii) Lemma 2, part (ii), implies that we can increase aggregate profit by creating a connected component from two non-singleton disconnected components (while possibly leaving some nodes isolated) with higher aggregate profit. Moreover, this connected component must be a nested split graph. Otherwise, there would exist a connected nested split graph with the same number of nodes and links which has a higher largest eigenvalue (Brualdi and Solheid, 1986), but the same number of links and hence the same cost. This shows that for $c > 1$ the efficient network has at most one non-singleton component which is a nested split graph. Next, we show that for $c < \frac{n}{2\sqrt{n-1}}$ the efficient graph is not empty. Note that aggregate profit of the empty graph is zero. Hence, the efficient graph G^* is not empty if we find a graph with positive total profits in that region of the cost. In particular, this holds for the star $K_{1,n-1}$. Aggregate profit of the star is given by $\Pi(K_{1,n-1}) = n\sqrt{n-1} - 2(n-1)c$ and therefore $\Pi(K_{1,n-1}) > 0$ whenever $n > 2c(c+\sqrt{c^2-1})$.[37] Solving for c, we obtain

[37] For instance, for $c = 1$ the condition holds for $n \geqslant 2$.

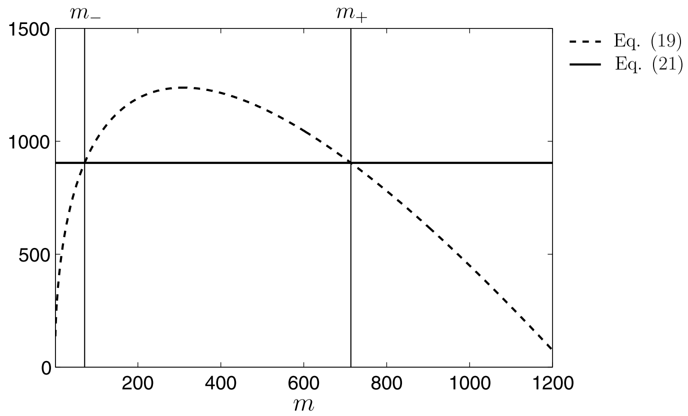


Fig. 5. The two bounds from Eqs. (19) and (21) for $n = 100$ and $c = 2$.

that the condition holds if $c < \frac{n}{2\sqrt{n-1}}$. Since the efficient graph is not empty, there must exist a non-singleton component. Putting the above results together shows that for $1 \leqslant c < \frac{n}{2\sqrt{n-1}}$ there exists a non-singleton component which must be a connected nested split graph.

(iv) For $c \geqslant \frac{n}{2\sqrt{n-1}}$, the proof of the previous item implies that the efficient graph G^* has at most one non-singleton component which is a nested split graph. However, this graph might also be empty.

(v) For $c > \sqrt{\frac{n(n-1)}{2}}$ we can show that the efficient graph must be empty. We first derive an upper bound for total profits of the efficient network G^*. Wilf (1967) has shown that, for any graph G, the largest eigenvalue $\lambda_{\mathrm{PF}}(G)$ is bounded from above by $\lambda_{\mathrm{PF}}(G) \leqslant \sqrt{2m(1-\frac{1}{n})}$. Therefore, aggregate profits for any graph G (including the efficient graph G^*) are bounded by

$$\Pi(G, c) \leqslant n\sqrt{2m\left(1 - \frac{1}{n}\right)} - 2mc. \tag{19}$$

Setting the RHS of Eq. (19) to zero gives the following bound for the maximum number of links in the efficient graph

$$m^* \leqslant \frac{n(n-1)}{2c^2}. \tag{20}$$

From Eq. (20) we see that m^* must be zero if $c > \sqrt{\frac{n(n-1)}{2}}$ and the efficient graph is the empty graph $\bar{K}_n$. Due to the monotonicity of the bound, $\bar{K}_n$ is also the unique efficient graph for $c > \sqrt{\frac{n(n-1)}{2}}$. □

Proof of Proposition 2. (i) From Proposition 1 we know that the efficient graph is complete, if $c \in [0, 0.5]$ so that $m^* = \binom{n}{2}$.

(ii) We have derived an upper bound for aggregate profits of the efficient graph G^* in Eq. (19) of the proof of Proposition 1. Note that the RHS of Eq. (19) is concave in the number of links m. Moreover, in Eq. (21) of the proof of Proposition 3 we derive a lower bound on aggregate profits of the graph $F_{n,d}$. This lower bound is also a lower bound on aggregate profits of the efficient graph G^*. The RHS of Eq. (21) is a convex function of n, which attains its maximum at the highest possible value of n. This maximum is independent of m. An illustration for $n = 100$ and $c = 2$ of the two bounds on aggregate profits in Eqs. (19) and (21) can be seen in Fig. 5. We consider only values of cost such that the bound in Eq. (21) is positive, which is equivalent to requiring that $c < c_1^* = \frac{n+2\sqrt{2n^2(n-1)}}{8n-9}$. Aggregate profits of the efficient graph G^* are positive in the region where the upper bound in Eq. (19) exceeds the lower bound in Eq. (21). Thus, by equating the RHS of Eq. (19) with Eq. (21) it follows that $m_-(n, c) \leqslant m^* \leqslant m_+(n, c)$, where the explicit expressions for $m_-(n, c)$ and $m_+(n, c)$ are given in the proposition. Incidentally, notice that both the derivatives of $m_-(n, c)$ and $m_+(n, c)$ with respect to c are negative as long as $c > \frac{n+3\sqrt{n(n-1)}}{8n-9}$, which approaches 0.5 for large n.

(iii) The desired bound for the maximum number of links in the efficient graph is given in Eq. (20) in the proof of Proposition 1. Note that for $c > c_1^* = \frac{n+2\sqrt{2n^2(n-1)}}{8n-9}$ the lower bound on aggregate profits in Eq. (21) is negative. This implies that the bound for m^* in Eq. (20) is lower than m_+ derived in part (ii) of the proof. The fact that the result holds for $c \leqslant c_2^*$ follows from item (iv).

(iv) From Eq. (20) it follows that for $c > c_2^* = \sqrt{\frac{n(n-1)}{2}}$ the upper bound on the number of links m^* becomes negative, so that the efficient graph must be empty. □

Proof of Proposition 3. In order to prove the claim, we need a lower bound for the total profit of $F_{n,d}$, as well as an upper bound for the total profit of the efficient graph G^*. We then show that, if one chooses d appropriately, the relative difference between the two bounds vanishes for large n.

Recall that $F_{n,d}$ is the graph obtained from a complete graph K_d of d nodes and $n-d$ isolated nodes by connecting each isolated node to one (and the same node) of K_d via one link. The number of links m in this graph is determined by the size d of the clique, $m(d)=\binom{d}{2}+(n-d)$. Since $F_{n,d}$ contains K_d as a subgraph, the largest real eigenvalue of $F_{n,d}$ is larger or equal to the one of K_d, which is $\lambda_{\mathrm{PF}}(K_d)=d-1$. Therefore, total profits of the graph $F_{n,d}$ are bounded from below as follows: $\Pi(F_{n,d})=n\lambda_{\mathrm{PF}}(F_{n,d})-2m(d)c\geqslant n(d-1)-2m(d)c$. Since this inequality is valid for any integer d, such that $1\leqslant d\leqslant n$, we are interested in the value of d such that $d=\arg\max_{1\leqslant k\leqslant n}\{n(k-1)-2m(k)c\}$, where $m(k)=\binom{k}{2}+(n-k)$ and $k\in\mathbb{N}$. By computing the first and second derivatives of the concave objective function $n(k-1)-2m(k)c$ with respect to k, one finds that its maximum occurs at $k=\frac{n+3c}{2c}$. Note that if $\frac{n+3c}{2c}>n$, then the maximum occurs at $k=n$. For our purposes, one can take d as the closest integer to this value. Notice that, as a consequence, d converges to $\frac{n}{2c}$ for large n. Replacing $d=\frac{n+3c}{2c}$ in the expression of $\Pi(F_{n,d})$, we obtain a lower bound, which is independent of d, and given by

$$\Pi(F_{n,d},c)\geqslant\frac{n^2+n(2c-8c^2)+9c^2}{4c}. \tag{21}$$

Next, note that the number of links maximizing the upper bound of Eq. (19) for $c\geqslant 1/2$ is given by $m=\frac{n(n-1)}{8c^2}$. For $c\leqslant 1/2$ this is $\frac{n(n-1)}{2}$, corresponding to the complete graph K_n. Inserting this specific value for the number of links into the upper bound on aggregate profits delivers an upper bound for the efficient graph G^*, which is independent of m, and given by

$$\Pi\big(G^*,c\big)\leqslant\frac{n(n-1)}{4c}.$$

We know that $\Pi(F_{n,d},c)\geqslant\frac{n^2+n(2c-8c^2)+9c^2}{4c}$. Therefore, denoting by $d^*=\arg\max_d\Pi(F_{n,d},c)$, we find that

$$\epsilon=\frac{\Pi(G^*,c)-\Pi(F_{n,d^*},c)}{\Pi(G^*,c)}\leqslant\frac{\Pi(G^*,c)-\Pi(F_{n,d^*},c)}{\Pi(F_{n,d^*},c)}\leqslant\frac{c^2(8n-9)-n-2nc}{n^2+2nc-c^2(8n-9)}.$$

For large n, the right-hand side of the above inequality converges to zero, and so does the relative error ϵ. □

Proof of Proposition 4. (vi) In a dominant-group architecture $D_{n,k}$ with $n\geqslant k+2$ there is at least a pair of isolated nodes. A link between two isolated nodes results in $\Delta\lambda_{\mathrm{PF}}=1$ and is thus profitable as long as $c<1$. Therefore, $D_{n,k}$, $n\geqslant k+2$, is not stable for $c\in[0,1)$. In general, it follows that any network with two or more isolated nodes is not stable in the cost range $c\in[0,1)$.

In the case of $n=k+1$, $D_{n,k}$ consists of a clique and an isolated node. The largest eigenvalue of the graph resulting from connecting the isolated node to one node in the clique is given by the largest root of the polynomial $\lambda^3+\lambda^2(n-2)-\lambda n+(n-2)$, which is given by $\lambda=\frac{1}{6}(4-2n+2^{4/3}(4-n+n^2)A^{-1}+2^{2/3}A)$, where $A=(70-33n+3n^2-2n^3+3\sqrt{516-492n+141n^2-42n^3+9n^4})^{1/3}$. From the expression for λ one finds that there exists a bound $c_1(n)=\lambda-(n-1)$ such that for $c>c_1(n)$ the creation of the link is not profitable. For $D_{n,k}$ to be stable, it must also hold that no link within the clique is removed. From the proof of Proposition 8 in König et al. (2011), this holds if and only if $c<c_2(n)=\frac{1}{2}(n+1-\sqrt{n^2+2n-7})$. Note that $c_2(n)<1$ for all n. There exist values of c such that $c_1(n)<c<c_2(n)<1$ and therefore in this range of cost $D_{n,k}$ is stable if $n=k+1$ but is not stable if $n\geqslant k+2$. We find that the range of suitable values of n decreases with the cost. For instance, for $c=0.05$ the structure is stable for $5<n<50$, while for $c=0.2$ it is only stable for $4\leqslant n\leqslant 10$.

(vii) This is shown in the first part of the proof of (vi). □

Proof of Proposition 5. (i) From Proposition 1, item (ii), we know that the complete graph is the unique efficient graph for $c\leqslant\frac{n}{2n-1}$. This bound is equivalent to $n\leqslant\infty$ for $0\leqslant c\leqslant 0.5$ and $n\leqslant\frac{c}{2c-1}$ for $0.5<c\leqslant 1$. This implies that the complete graph is uniquely efficient in the region to the left of the dotted curve $g(c)=\frac{c}{2c-1}$ in Fig. 4. On the other hand, Proposition 8 in König et al. (2011) implies that the complete graph is not stable for $n>\lfloor\frac{2-c(1-c)}{c}\rfloor$, that is, above the curve $f_1(c)$ in Fig. 4. It follows that the unique efficient graph is not stable if $c\leqslant 0.5$ and $n>\frac{2-c(1-c)}{c}$, or if $c>0.5$ and $\frac{2-c(1-c)}{c}<n\leqslant\frac{c}{2c-1}$. In Fig. 4 this corresponds to the region above curve $f_1(c)$ and to the left of curve $g(c)$.

(ii) Let G be a connected nested split graph and denote by $\Delta\lambda_{\mathrm{PF}}(G)=\lambda_{\mathrm{PF}}(G)-\lambda_{\mathrm{PF}}(G-ij)$ for a link ij in G. Further, let $\mathbf{v}$ be the eigenvector associated with $\lambda_{\mathrm{PF}}(G)$, normalized such that $\mathbf{v}^\top\mathbf{v}=1$. Then we have that

$$\Delta\lambda_{\mathrm{PF}}(G)=\mathbf{v}^\top\big(\lambda_{\mathrm{PF}}(G)-\lambda_{\mathrm{PF}}(G-ij)\big)\mathbf{v}\leqslant\mathbf{v}^\top\big(\mathbf{A}(G)-\mathbf{A}(G-ij)\big)\mathbf{v}=2v_iv_j,$$

where we have used the fact that $\lambda_{\mathrm{PF}}(G-ij)>\mathbf{v}^\top\mathbf{A}(G-ij)\mathbf{v}$ as long as $\mathbf{v}$ is not an eigenvector of $\mathbf{A}(G-ij)$ with eigenvalue $\lambda_{\mathrm{PF}}(G-ij)$, as follows from the Rayleigh–Ritz theorem (Horn and Johnson, 1990). It is enough to prove that there exists at least one link that is beneficial to remove. Choosing i such that $v_i = v_{\max}$ and j such that $v_j = v_{\min}$ (where $v_{\max}$ and $v_{\min}$ denote the largest and smallest eigenvector components, respectively), we obtain $\Delta\lambda_{\mathrm{PF}}(G) \leqslant 2v_{\max}v_{\min}$.

Recall that Δ and δ denote the maximum and minimum degrees in G. The nested neighborhood structure of G implies that the node with maximal (minimal) eigenvector centrality has also maximal (minimal) degree: $d_i = \Delta$ and $d_j = \delta$. Using the fact that in a connected nested split graph $v_{\min} = \frac{1}{\lambda_{\mathrm{PF}}(G)}\sum_{k\in N_j} v_k \leqslant \frac{\delta}{\lambda_{\mathrm{PF}}(G)}v_{\max}$, we obtain $\Delta\lambda_{\mathrm{PF}}(G) \leqslant 2v_{\max}^2\frac{\delta}{\lambda_{\mathrm{PF}}(G)}$. Moreover, Papendieck and Recht (2000) have shown that $v_{\max} \leqslant 1/\sqrt{2}$, so that we obtain

$$\Delta\lambda_{\mathrm{PF}}(G) \leqslant \frac{\delta}{\lambda_{\mathrm{PF}}(G)}, \tag{22}$$

which converges to zero as n becomes large (for a fixed value of δ) since $\lambda_{\mathrm{PF}}(G)$ increases with n for a connected nested split graph.

For larger values of δ we use the fact that $v_i \leqslant 1/\sqrt{1+\frac{\lambda_{\mathrm{PF}}(G)^2}{d_i}}$.[38] In our case this implies that

$$\Delta\lambda_{\mathrm{PF}}(G) \leqslant \frac{2}{\sqrt{(1+\frac{\lambda_{\mathrm{PF}}(G)^2}{\Delta})(1+\frac{\lambda_{\mathrm{PF}}(G)^2}{\delta})}}. \tag{23}$$

Since G is connected and contains a spanning star as an induced subgraph, we have that $\Delta = n-1$.

Next, we observe that if the minimum degree in G is δ, then the stepwise property of the adjacency matrix of G implies that G contains δ spanning stars (with distinct centers) as subgraphs. Let us denote this graph by $S_{n,\delta} \subset G$. The adjacency matrix for a graph $S_{5,2}$ ($\delta = 2$) is shown below:

$$\mathbf{A}(S_{5,2}) = \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{pmatrix}.$$

A similar calculation as the one in the proof of Proposition 8 in König et al. (2011) shows that the largest eigenvalue of $S_{n,\delta}$ is given by $\lambda_{\mathrm{PF}}(S_{n,\delta}) = \frac{1}{2}(n-1-\delta+\sqrt{(n-1)^2+2\delta(n+1)-3\delta^2})$. Inserting the above expression into Eq. (23) shows that in the limit of large n we have that $\Delta\lambda_{\mathrm{PF}}(G) \leqslant \frac{2}{\delta+1}$. For the case of $\delta = 2$ Eq. (22) shows that the change in eigenvalue converges to zero as n becomes large. Hence, we need to consider only $\delta \geqslant 3$, where it follows that $\Delta\lambda_{\mathrm{PF}}(G) \leqslant 0.5$.

(iii) From Eq. (21) we know that $\Pi(F_{n,d}, c) \geqslant \frac{n^2+n(2c-8c^2)+9c^2}{4c}$, which is a lower bound for aggregate profits of the efficient graph G^*. We have that $\Pi(F_{n,d}, c) > 0$ – which implies that the efficient graph is non-empty – as long as $c < \frac{n+2\sqrt{2n^2(n-1)}}{8n-9}$. However, for $c > 1$ the empty graph is the unique stable network. Hence, the efficient network G^* is not stable if $1 < c < \frac{n+2\sqrt{2n^2(n-1)}}{8n-9}$. Note that the upper bound in the last inequality is larger than 1 if $n > 3$. □

References

Ahuja, G., 2000. Collaboration networks, structural holes, and innovation: A longitudinal study. Admin. Sci. Quart. 45, 425–455.

Aouchiche, M., Bell, F., Cvetkovic, D., Hansen, P., Rowlinson, P., Simic, S., Stevanovic, D., 2008. Variable neighborhood search for extremal graphs. 16. Some conjectures related to the largest eigenvalue of a graph. Europ. J. Oper. Res. 191 (3), 661–676.

Bala, V., Goyal, S., 2000. A noncooperative model of network formation. Econometrica 68 (5), 1181–1230.

Ballester, C., Calvó-Armengol, A., Zenou, Y., 2006. Who's who in networks. Wanted: The key player. Econometrica 74 (5), 1403–1417.

Bell, F.K., 1991. On the maximal index of connected graphs. Linear Algebra Appl. 144, 135–151.

Bell, F.K., 1992. A note on the irregularity of graphs. Linear Algebra Appl. 161, 45–54.

Bloch, F., Jackson, M., 2006. Definitions of equilibrium in network formation games. Int. J. Game Theory 34 (3), 305–318.

Borgatti, S., Everett, M., 2000. Models of core/periphery structures. Soc. Netw. 21 (4), 375–395.

Bramoullé, Y., 2007. Anti-coordination and social interactions. Games Econ. Behav. 58 (1), 30–49.

Brualdi, R.A., Solheid, E.S., 1986. On the spectral radius of connected graphs. Publ. Inst. Math. 53, 45–54.

Burt, R., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge, MA.

Calvó-Armengol, A., Ilkilic, R., 2009. Pairwise-stability and Nash equilibria in network formation. Int. J. Game Theory 38, 51–79.

Carayol, N., Roux, P., 2009. Knowledge flows and the geography of networks: A strategic model of small world formation. J. Econ. Behav. Organ. 71 (2), 414–427.

Cioaba, S., Gregory, D., 2007. Principal eigenvectors of irregular graphs. Electron. J. Linear Algebra 16, 366–379.

Cohen, W., Levinthal, D., 1989. Innovation and learning: The two faces of R&D. Econ. J. 99, 569–596.

Copeland, A., Fixler, D., February 2009. Measuring the price of research and development output. BEA Working Paper No. 0044.

Corbo, J., Calvó-Armengol, A., Parkes, D., 2006. A study of Nash equilibrium in contribution games for peer-to-peer networks. SIGOPS Operation Systems Rev. 40 (3), 61–66.

Cvetkovic, D., Rowlinson, P., 1990. The largest eigenvalue of a graph: A survey. Linear Multilinear Algebra 28 (1), 3–33.

Cvetkovic, D., Rowlinson, P., Simic, S., 1997. Eigenspaces of Graphs. Encyclopedia Math. Appl., vol. 66. Cambridge University Press.

[38] See Theorem 3.2 in Cioaba and Gregory (2007).

Deroian, F., 2008. Dissemination of spillovers in cost-reducing alliances. Res. Econ. 62 (1), 34–44.
DiMasi, J., Hansen, R., Grabowski, H., 2003. The price of innovation: New estimates of drug development costs. J. Health Econ. 22 (2), 151–185.
Dosi, G., 1988. Sources, procedures and microeconomic effects of innovation. J. Econ. Lit. 26 (3), 1120–1171.
Galeotti, A., Goyal, S., 2009. Influencing the influencers: A theory of strategic diffusion. RAND J. Econ. 40 (3), 509–532.
Goyal, S., 2009. Connections: An Introduction to the Economics of Networks. Princeton University Press.
Goyal, S., Joshi, S., 2003. Networks of collaboration in oligopoly. Games Econ. Behav. 43 (1), 57–85.
Goyal, S., Joshi, S., 2006. Unequal connections. Int. J. Game Theory 34 (3), 319–349.
Goyal, S., Moraga-Gonzalez, J.L., 2001. R&D networks. RAND J. Econ. 32 (4), 686–707.
Hagberg, J., 2004. On degree variance in random graphs. PhD thesis, Stockholms Universitet, Statistiska Institutionen.
Hagedoorn, J., 2002. Inter-firm R&D partnerships: An overview of major trends and patterns since 1960. Res. Pol. 31 (4), 477–492.
Hellmann, T., Buechel, B., 2009. Under-connected and over-connected networks. FEEM Working Paper No. 38.
Hojman, D., Szeidl, A., 2008. Core and periphery in networks. J. Econ. Theory 139 (1), 295–309.
Horn, R.A., Johnson, C.R., 1990. Matrix Analysis. Cambridge University Press.
Jackson, M., 2008. Social and Economic Networks. Princeton University Press.
Jackson, M.O., Wolinsky, A., 1996. A strategic model of social and economic networks. J. Econ. Theory 71 (1), 44–74.
Jones, C.I., 1995. R&D-based models of economic growth. J. Polit. Economy 103, 759–784.
König, M.D., Battiston, S., Napoletano, M., Schweitzer, F., 2008. On algebraic graph theory and the dynamics of innovation networks. Netw. Heterog. Med. 3 (2), 201–219.
König, M.D., Battiston, S., Napoletano, M., Schweitzer, F., 2011. Recombinant knowledge and the evolution of innovation networks. J. Econ. Behav. Organ. 79 (3), 145–164.
Kogut, B., Zander, U., 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. Organ. Sci. 3 (3), 383–397.
Mahadev, N., Peled, U., 1995. Threshold Graphs and Related Topics. North-Holland.
Papendieck, B., Recht, P., 2000. On maximal entries in the principal eigenvector of graphs. Linear Algebra Appl. 310 (1–3), 129–138.
Persitz, D., 2009. Power in the heterogeneous connections model: The emergence of core–periphery networks. Fondazione Eni Enrico Mattei Working Paper 303.
Powell, W.W., Koput, K.W., Smith-Doerr, L., 1996. Interorganizational collaboration and the locus of innovation: Networks of learning in biotechnology. Admin. Sci. Quart. 41 (1), 116–145.
Powell, W.W., White, D.R., Koput, K.W., Owen-Smith, J., 2005. Network dynamics and field evolution: The growth of interorganizational collaboration in the life sciences. Amer. J. Econ. Sociology 110 (4), 1132–1205.
Seneta, E., 2006. Non-negative Matrices and Markov Chains. Springer.
Spence, M., 1984. Cost reduction, competition, and industry performance. Econometrica, 101–121.
Stevanovic, D., 2007. Research problems from the Aveiro workshop on graph spectra. Linear Algebra Appl. 423, 172–181.
Vega-Redondo, F., 2007. Complex Social Networks. Econom. Soc. Monogr. Cambridge University Press.
Veugelers, R., 1997. Internal R&D expenditures and external technology sourcing. Res. Pol. 26 (3), 303–315.
Weitzman, M.L., 1998. Recombinant growth. Quart. J. Econ. 113 (2), 331–360.
Westbrock, B., 2010. Natural concentration in industrial research collaboration. RAND J. Econ. 41 (2), 351–371.
Wilf, H.S., 1967. The eigenvalues of a graph and its chromatic number. J. London Math. Soc. 42, 330–332.