Collaboration in social networks

Luca Dall'Asta^{a,b,c,1}, Matteo Marsili^b, and Paolo Pin^d

^aDipartimento di Fisica, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129 Torino, Italy; ^bThe Abdus Salam International Center for Theoretical Physics, Strada Costiera 11, 34014 Trieste, Italy; ^cCollegio Carlo Alberto, Via Real Collegio 30, 10024 Moncalieri (Torino), Italy; and ^dDipartimento di Economia Politica, Università degli Studi di Siena, Piazza San Francesco 7, 53100 Siena, Italy

Edited* by Giorgio Parisi, University of Rome, Rome, Italy, and approved January 12, 2012 (received for review April 12, 2011)

The very notion of social network implies that linked individuals interact repeatedly with each other. This notion allows them not only to learn successful strategies and adapt to them, but also to condition their own behavior on the behavior of others, in a strategic forward looking manner. Game theory of repeated games shows that these circumstances are conducive to the emergence of collaboration in simple games of two players. We investigate the extension of this concept to the case where players are engaged in a local contribution game and show that rationality and credibility of threats identify a class of Nash equilibria—that we call “collaborative equilibria”—that have a precise interpretation in terms of subgraphs of the social network. For large network games, the number of such equilibria is exponentially large in the number of players. When incentives to defect are small, equilibria are supported by local structures whereas when incentives exceed a threshold they acquire a nonlocal nature, which requires a “critical mass” of more than a given fraction of the players to collaborate. Therefore, when incentives are high, an individual deviation typically causes the collapse of collaboration across the whole system. At the same time, higher incentives to defect typically support equilibria with a higher density of collaborators. The resulting picture conforms with several results in sociology and in the experimental literature on game theory, such as the prevalence of collaboration in denser groups and in the structural hubs of sparse networks.

The social network influences and constrains in nontrivial ways the behavior of individuals (1) but also contributes to aspects generically referred to as *social capital* (2–4), which favor the emergence of coordinated actions or collaboration[†].

The theoretical investigation about the emergence of collaboration and the fate of repeated actions on networks has, up to now, mostly focused on adaptive learning, imitation, and evolutionary game theory. The corresponding theoretical models (6–9) seem to suggest that sustaining collaboration is easier on sparse networks and spatial structures than on dense groups. On the contrary, there is no clear experimental evidence of such an effect. Results of recent experiments on voluntary contribution suggest instead that the establishment and maintenance of collaboration is easier on well connected groups than on sparse graphs (10–12), in some ways confirming the intuition behind the classical Coleman’s closure argument (13). The picture becomes more complex as we move to analyze individual positions in the network: Experiments performed on star-shaped groups suggest that central nodes are fostered to collaborate more than peripherals (12), corroborating another well known idea; i.e., Burt’s argument on the importance of structural holes (14).

However, the network is not only the channel for the achievement of payoffs or the exchange of information, which are the key ingredients of learning and imitative behavior. The network also supports the establishment of trust, norms, contracts, and other continuative collaborative relations in which the temporal dimension is crucial. In fact, a key aspect of a social network is that individuals connected by a link interact repeatedly with each other, calling into play forward looking strategic behavior typical of the theory of repeated games. This feature provides a formal framework for concepts such as threats, punishment, and credibility, which have been so enlightening on the emergence of collaboration in simple setups (e.g., the prisoner’s dilemma) (15).

Here we investigate the extension of the theory of repeated games to network games. We focus on simple local contribution games which provide a perfect framework to highlight the main conceptual issues, and are closely linked to experimental works on voluntary contribution (10–12, 16). Under very general assumption on the structure of the network, we prove that, in a multiplayer setting, rationality is consistent with conditional collaboration that is (i) player-specific and (ii) reciprocal. Interestingly, the fact that collaboration can be sustained through a reciprocal relation of control and punishment is a largely verified experimental fact (16).

We satisfy these requirements with a refinement of Nash equilibria, that we call Collaborative Equilibria, with a precise graph-theoretical interpretation. This refinement turns the problem of finding and characterizing Nash equilibria into that of finding specific subgraphs of the social network, which can be addressed by graph theoretical methods (17, 18). These techniques provide considerable insight on the nature of possible collaborative structures which can be supported, depending on incentives and on the topology of the social network. In particular, we find that when the contribution cost is low, collaboration can be sustained by local commitments (corresponding to dimers or loops on the network). But when costs are high, collaboration requires non-local structures which span a finite fraction of the system. This structure has clear implications in terms of systemic fragility, because when costs are high a single individual defection may bring to the collapse of the whole collaborative network.

Our analysis of Collaborative Equilibria on ensembles of random graphs unveils a generic picture which, as will be discussed in the closing section, has many correspondences with well known results in sociology and in experimental economics.

A Repeated Game of Local Contribution

We focus on a class of local contribution games, recently popularized in refs. 16, 19, which have the peculiarity that only the neighbors of a contributor and not the contributor herself benefit of the positive effects of a contribution. In this regard, we can interpret these local contribution games as a simple extension of the celebrated prisoners’ dilemma to the case where there are more than two players, where the range of possible interactions between individuals is limited only to local ones by a fixed network structure, and where agents play the same strategy against each neighbor.

Author contributions: L.D., M.M., and P.P. designed research; L.D., M.M., and P.P. performed research; and L.D., M.M., and P.P. wrote the paper.

The authors declare no conflict of interest.

*This Direct Submission article had a prearranged editor.

[†]We will explicitly refer to this endogenous profitable mutual exchange as *collaboration*, and not as *cooperation*, to stress the fact that it emerges from individual opportunistic incentives and not from group-based profit maximization, as in the cooperative games already introduced in (5).

[‡]We adopt the notation $(N, \vec{N})$ to specify the social network, where N is the set of nodes and the i th element of $\vec{N}$, denoted N_i , is the subset of the neighbors of i ($i \in N_i$). The social network is undirected: $i \in N_j$ implies $j \in N_i$.

[§]To whom correspondence should be addressed. E-mail: luca.dallasta@polito.it.

This article contains supporting information online at www.pnas.org/lookup/suppl/doi:10.1073/pnas.1105757109/-DCSupplemental.

The Model. Players occupy the nodes N of an undirected social network and they interact with their neighbors in a local contribution game. Specifically, each player $i \in N$ has the option of either contributing ($s_i = 1$) or not ($s_i = 0$) to a local public good that has effect only on the neighbors of i in N_i^* . Contributing is costly, which means that players who contribute incur a cost $X_i > 0$. At the same time, each player receives a unitary payoff from all agents $j \in N_i$ who contribute in her neighborhood. The payoff function of agent i is then given by

$$\pi_i(s_i, s_{-i}) = -X_i s_i + \sum_{j \in N_i} s_j, \quad [1]$$

where s_{-i} stands for the vector of choices of all other players, except i . It is clear that, in the single stage game, a positive s_i is only a cost for player i , and hence defection (i.e., $s_i = 0$) is a dominant strategy for all $X_i > 0$: $\pi_i(0, s_{-i}) > \pi_i(1, s_{-i})$. If the game is played just once, the unique Nash equilibrium is one where $s_i = 0$ for all $i \in N$. When players are engaged in repeated plays of the game, with the same opponents, a much richer set of outcomes is possible. Indeed, a continuum set of outcomes can be sustained as Nash equilibria (15).

Let the game above be played repeatedly at discrete times $t = 0, 1, 2, 3, \dots$ and let $s_i^{(t)}$ be the option taken by agent i at time t . A strategy now becomes an *action plan* $\sigma_i = \{s_i^{(t)}, t = 0, 1, 2, \dots\}$, which specifies the behavior of agent i at all times. In particular, $s_i^{(t)}$ may be any function of the opponents' behavior $s_{-i}^{(t')}$ in the previous stage games $t' < t$. The payoff function is generalized to an intertemporal utility function

$$\pi_i(\sigma_i, \sigma_{-i}) = (1 - \delta) \sum_{t \geq 0} \pi(s_i^{(t)}, s_{-i}^{(t)}) \delta^t, \quad [2]$$

where $\delta \in (0, 1]$ is the factor by which agents discount future payoffs with respect to present ones. For $\delta \rightarrow 0$ the game reverts to the single stage game with the unique noncollaborative equilibrium. The limit $\delta \rightarrow 1$ of (temporally) very far-sighted players is instead much more interesting, because players can choose from a huge (a priori infinite) space of possible strategies.

Trigger Strategies. To attack this problem, we start considering the simpler and well known situation of a two-player game, where enforceable payoffs can be attained as Nash equilibria by reducing the set of possible strategies to a particular subset, known as *trigger strategies* (20). Trigger strategies encode the idea of *punishment*: agent i collaborates a priori, but if her opponent j misbehaves (i.e., $s_j^{(t)} = 0$ for some $t \geq 0$), she will punish her with defective behavior for the infinite future (i.e., $s_i^{(t')} = 0$ for all $t' > t$). If the opponent is "threatened" in this way, her best reply, if δ is large enough and $X_i < 1$, is to collaborate, or equivalently to use the same trigger strategy. This strategy in fact guarantees the collaboration payoff $1 - X_i > 0$ in each round, as opposed to the zero payoff she would get if both defected. A crucial issue is that *threats of punishment must be credible*: a rational opponent will not consider credible a trigger strategy that inflicts a payoff loss to the player herself.

How does this insight carry over to multiplayer settings, where players interact with their neighbors on the network? It is instructive to consider the case of three players, without loss of generality call them 1, 2, and 3, all connected together in a closed triangle[§], with $X_i < 1$. Imagine that players 1 and 2 are both using trigger strategies, making their collaboration conditional on the collaboration of the others. How should player 3 behave? If she takes the threat of other players seriously, then she will also collaborate. But should she consider their threat credible? No, because if player 3 were to defect, and both 1 and 2 also turn

to defection, they would get a payoff of zero, whereas if they were to continue collaborating they would get a payoff of $1 - X_i > 0$ in each stage game. Hence the threat of players 1 and 2 is not credible and player 3 better defects, free-riding with a payoff of 2. On the contrary, if we assume $1 < X_i < 2 \forall i$ then full collaboration is individually rational and can be sustained in the repeated game as a Nash equilibrium outcome by means of credible threats. This example points out that generalizing trigger strategies entails making the strategies conditional on the behavior of a subset of the players: for $X_i < 1$, player 1 will punish player 2 but not player 3, because only one collaborator is needed in the neighborhood to make collaboration the best choice. In other words, *control by punishment is player-specific*^{||}.

Also, the same arrangement of players shows that *punishment should be reciprocal*. Imagine a situation where player 1 punishes 2, 2 punishes 3, and 3 punishes 1. Should 3 abide to this arrangement? No. Again if she decides to defect, she may argue that it is not credible that 2 will punish her, because again 2 would end up with a lower payoff (zero) than what she would get by continuing collaborating with 1.

Collaborative Equilibria. The three conditions that punishment must be: (i) *credible*, (ii) *player-specific*, and (iii) *reciprocal* identify a notion of Nash equilibrium sustained by trigger strategies on general graphs. We shall first consider the case of trees (Proposition 1) and then generalize to general graphs (Proposition 2). Central to our analysis is the notion of *nearsightedness*; i.e., we focus on Nash equilibria where players can only punish their neighbors on the graph.

Let $\sigma = 0$ be the "always defect" strategy and $\tau(\Delta)$ be a trigger strategy, conditional on the behavior of agents $j \in \Delta \subseteq N$. An agent playing $\tau(\Delta)$ will collaborate as long as all agents in Δ also collaborate. Let $C \subseteq N$ be the subset of agents who play trigger strategies, whereas agents $i \notin C$ always defect ($\sigma = 0$). Individuals only control their neighbors; i.e., if $i \in C$ plays $\tau(\Delta_i)$, then $\Delta_i \subseteq N_i$. It is convenient to introduce also the set of *direct punishers* of $i \in C$, which is $\Gamma_i = \{j: i \in \Delta_j\} \subseteq C \cap N_i$. Then let $c_i = |C \cap N_i|$ be the number of collaborators in i 's neighborhood and $\gamma_i = |\Gamma_i|$ be the number of punishers of i .

We start with a result on networks without loops; i.e., trees.

Proposition 1 (Collaborative Equilibria in Trees) Assume that the network has no loops. The arrangement where, for all $i \in C$, $X_i \leq \gamma_i < X_i + 1$, is a Nash equilibrium of the repeated game where $\Gamma_i = \Delta_i$ (i.e., $j \in \Gamma_i$ if and only if $i \in \Gamma_j$), for a sufficiently large discount factor δ

We refer the interested reader to [SI Text](#) for a detailed proof. We shall call Nash equilibria of this type *Collaborative Equilibria* (CE). In a CE, the defection of a player will generate a cascade of defections due to punishment. When player i considers whether she should punish defection of $j \in N_i$, these indirect effects need to be considered. However, on trees, the defection of a neighbor $j \in N_i$ of player i , will only have consequences on the subtree downstream of j , and will have no direct effects on other neighbors of i . The proof of Proposition 1 relies on this fact.

The above proposition proves only sufficiency, as there could be other trigger strategies equilibria; e.g., the one in which there is no nearsightedness and in equilibrium any collaborator would immediately defect as a result of any defection. Interestingly, CE turn out to correspond, on trees, to subgame perfect Nash equilibria of a similar game with the same payoffs (22, 23). In this related game, the assumptions in Proposition 1 are also necessary, which provides even more support for the notion of CE as a re-

[§]From now on, for the sake of simplicity and without loss of generality, we focus on the limit $\delta \rightarrow 1$ where the utility is dominated by the asymptotic behavior of players. Each statement derived for $\delta \rightarrow 1$ will be true for a $\delta < 1$, which is large enough.

^{||}This example is discussed also in the [SI Text](#).

^{||}This notion was already implicitly present in several multiplayer generalization of the Prisoner's Dilemma introduced to study the free-riding problem in groups of individuals (21).

finement of repeated games on networks. We refer to the [SI Text](#) for a detailed treatment of this issue.

On general networks, the indirect effects of cascades of punishment need to be taken into account. Consider a strategy profile as in Proposition 1, where for all $i \in C$, $X_i \leq \gamma_i < X_i + 1$ and $\Gamma_i = \Delta_i$. Let $D_{i \rightarrow i}$ be the neighbors of i , who are not in Γ_i and who would defect as an indirect consequence of the defection of $j \in C$, if i does not defect. Let

$$D_{i \rightarrow i} = \bigcup_{j \in \Gamma_i} D_{j \rightarrow i}$$

be the set of *indirect punishers* of i ; i.e., all those neighbors not in Γ_i that would defect as an indirect consequence of the defection of i herself. Then we have the following result.

Proposition 2. (Collaborative Equilibria) Consider the arrangement where, for all $i \in C$, $X_i \leq \gamma_i < X_i + 1$ and $\Gamma_i = \Delta_i$. If moreover, for all $i \in C$ and $j \in \Gamma_i$, we have that $D_{j \rightarrow i} = D_{i \rightarrow i}$, then this arrangement is a Nash equilibrium of the repeated game, for a sufficiently large discount factor δ

We refer the interested reader to [SI Text](#) for a detailed proof. A direct consequence of Proposition 2 is that it turns the characterization of Nash equilibria into a graph theoretical problem. Whenever the conditions of Propositions 2 applies, every collaborative equilibrium of the local contribution game identifies a subgraph $(C, \bar{\Gamma})$ of the social network (N, N) , with $C \subseteq N$, $\Gamma_i \subseteq N_i$, and $|\Gamma_i|$ being the smallest integer larger than X_i . Every subgraph $(C, \bar{\Gamma})$ with these properties supports a Nash equilibrium where players $i \in C$ collaborate ($s_i^{(t)} = 1$) and players not in C defect ($s_i^{(t)} = 0$). Every player $i \in C$ belongs to one and only one component in $(C, \bar{\Gamma})$.

The technical condition $D_{j \rightarrow i} = D_{i \rightarrow i} \forall i \in C$ and $j \in \Gamma_i$, seems more restrictive than it actually is. It is clearly satisfied on trees, where $D_{j \rightarrow i} = D_{i \rightarrow i} = \emptyset$, and is expected to be verified on large random graphs, which typically have a tree-like structure (18), when C is the union of connected components of few sites (as in the example below with $\gamma_i = 1 \forall i$). It is also true on the polar case of the complete graph, where again $D_{i \rightarrow i} = D_{j \rightarrow i} = \emptyset$. Indeed, in a general graph, imagine two collaborating neighbors of i , call them j and k , and suppose that $k \in D_{i \rightarrow i}$ but $k \notin D_{j \rightarrow i}$: this can happen only if j and k have no other paths connecting them in the subgraph $(C, \bar{\Gamma})$ apart from the one passing through i . Interestingly, in contrast with the intuition from Proposition 1, the emergence of loops and cliques makes the statement of Proposition 2 very likely to hold in dense random networks. At the same time, on sparse graphs, the network is tree-like and it is often the case that the subgraph $(C, \bar{\Gamma})$ has no loops. Then, in this case, Proposition 1 applies.

For this reason, an efficient strategy to find Collaborative Equilibria is to first solve the graph theoretical problem of identifying subgraphs $(C, \bar{\Gamma})$ that satisfy the conditions of Proposition 1 and then to check that the technical condition in Proposition 2 is satisfied. This is the strategy we will follow in the sequel to investigate the structure of collaborative equilibria and their computational complexity.

The Complexity of Collaboration on Networks

How many collaborative equilibria exist on a given social network and how does this number depend on the number of players, the payoffs, and the network structure? How hard is it to compute an equilibrium? How does the equilibrium “respond” to local perturbations? The mapping of collaborative equilibria to collections of possibly disconnected subgraphs of a given socio-economic network allows us to address a number of questions. Some of these questions can be answered in very general terms: For example, if $0 < X_i < 1 \forall i$, the subgraphs which support collaborative equilibria are collections of disjoint “dimers” (see

Fig. 1 left). The number of dimer covers is generically expected to increase exponentially with the size of the network. Local deviations have only local effects, as the defection of one player affects at most the behavior of the other player on the same dimer. If $1 < X_i < 2$ collaborative equilibria coincide with configurations of loops on the social network and larger X_i entail more complex structures (see Fig. 1 right). Correspondingly, the answer to the questions above becomes nontrivial. Also, the problem of finding subgraphs of a given network can be computationally very hard^{**}, but it is typically very easy for dimers and loops, or for specific classes of graphs, such as planar graphs. Note that Proposition 2 applies for dimers and loops because $D_{j \rightarrow i} = \emptyset$ for all i, j in the first case, whereas $D_{j \rightarrow i} = D_{i \rightarrow i}$ on loops.

Collaborative Equilibria on Random Graphs. Quantitative predictions are possible for networks which are drawn from ensembles of random graphs. For large network sizes ($|N| \rightarrow \infty$) several properties attain a limit which is independent of the particular realization of the social network. We refer to these as *typical* properties; i.e., properties which are expected to hold with high probability for large networks. The local tree-like structure of large graphs with a finite degree makes the characterization of typical behavior tractable by means of message passing techniques (18), such as Belief Propagation (BP), which are exact on trees and are approximately correct also for finite, moderately clustered graphs. The theoretical treatment is formally the same as in refs. 25–27, which have considered similar graph-theoretic problems^{††}. BP allows one to derive results on ensembles of graphs in the limit $|N| \rightarrow \infty$ (18) but it also provides efficient heuristic algorithms to find collaborative equilibria on given graph instances (28, 29). The messages which are exchanged in the BP algorithm are the probabilities $\mu_{i \rightarrow j} = P\{i \in \Gamma_j \cap C\}$ that player i collaborates and punishes j , in the collaborative equilibrium^{‡‡}. These probabilities, are updated through the equation (26)

$$\mu_{i \rightarrow j} = \frac{e^{-\epsilon} Z_{N_j \setminus j \rightarrow i}^{\gamma_i - 1}}{Z_{N_j \setminus j \rightarrow i}^0 + e^{-\epsilon} Z_{N_j \setminus j \rightarrow i}^{\gamma_i - 1} + e^{-\epsilon} Z_{N_j \setminus j \rightarrow i}^{\gamma_i}}, \quad [3]$$

where, for any integer q and subset $V \subseteq N$,

$$Z_{V \rightarrow i}^q = \sum_{U \subseteq V} \mathbb{I}_{|U|=q} \prod_{j \in U} \mu_{j \rightarrow i} \prod_{k \in V \setminus U} (1 - \mu_{k \rightarrow i}) \quad [4]$$

and the indicator function $\mathbb{I}_{|U|=q}$ restricts the sum only to subsets of q elements. In words, the numerator in the right hand side (r.h.s.) of Eq. 3, which is based on Eq. 4, asserts that i should control j if there are $\gamma_i - 1$ other players $k \neq j$ who control i . The denominator expresses the fact that, in a collaborative equilibrium, three situations are possible: (i) $i \notin C$, and hence i need not be controlled by any neighbor, (ii) $i \in \Gamma_j$ and $i \in C$, or (iii) $i \notin \Gamma_j$ and $i \in C$; i.e., i is already controlled by γ_i neighbors and does not need to control j . The parameter $\epsilon \in (-\infty, +\infty)$ is a statistical weight for collaborators, which is introduced to bias the distribution over equilibria towards those with a higher ($\epsilon < 0$) or lower ($\epsilon > 0$) density of collaborators. The probability $P\{i \in C\}$ that a player i collaborates or the probability $P\{i \in \Gamma_j\}$ that players $i, j \in C$ conditionally collaborate can be expressed in terms of the solution $\{\mu_{i \rightarrow j}\}$ of the set of Eqs. 3 and 4 as follows

^{**}For instance, in the case of random regular subgraphs; i.e., $\gamma_i = \gamma \forall i$, the problem is known to be NP-complete on general graphs (24).

^{††}A finer approximation, going beyond the so-called replica-symmetric ansatz, is possible (17). This approximation allows one to explore possible organization of the space of solutions which are more complex than the one described by a single Gibbs measure, as in the BP approach, but it goes beyond the scope of the present paper and will not be pursued further.

^{‡‡}In the subgraph problem, this is the probability that the link (i, j) is part of the subgraph, in the modified graph in which the link from i to j is removed. On trees, this operation disconnects the graph in subtrees.

Fig. 1. Example of collaborative equilibria for $N = 40$ players on a regular random graph ($|N_i| = K = 4$) with $0 < X_i < 1, \forall i$ (left) and $2 < X_i < 3, \forall i$ (right). Defectors are in black whereas collaborators are in yellow. Thick links are those between players who collaborate conditionally on each other ($i \in \Gamma_j$ and $j \in \Gamma_i$).

$$P\{i \in C\} = \frac{e^{-\epsilon} Z_{N_i \rightarrow i}^{\gamma_i}}{Z_{N_i \rightarrow i}^0 + e^{-\epsilon} Z_{N_i \rightarrow i}^{\gamma_i}} \quad [5]$$

$$P\{i \in \Gamma_j\} = \frac{\mu_{i \rightarrow j} \mu_{j \rightarrow i}}{\mu_{i \rightarrow j} \mu_{j \rightarrow i} + (1 - \mu_{i \rightarrow j})(1 - \mu_{j \rightarrow i})}. \quad [6]$$

In practice, Eqs. 3 and 4 can be iterated on a specific graph, substituting the value $\mu_{i \rightarrow j}$ with the value of the function on the r.h.s., until a fixed point is reached, which is guaranteed to occur on tree-like structures (18). The fixed point, however, is not unique. Indeed, for any collaborative equilibrium $(C, \bar{\Gamma})$, binary messages $\mu_{i \rightarrow j} = \mathbb{1}_{i \in \Gamma_j \cap C} \forall i, j$ are a solution of the BP equations. These “pure strategy” fixed points coexist with an internal solution⁸⁸. As in similar problems (27), the BP iteration converges to the internal solution, and makes possible a series of estimates on the statistical properties of the problem. In particular, we can compute the number $\mathcal{N}(\rho)$ of collaborative equilibria as a function of the density of collaborators $\rho = |C|/|N| \simeq \sum_{i \in N} P\{i \in C\}/|N|$. For large population size $|N|$, the number of equilibria is, to leading order, exponentially large in $|N|$, and one can define the entropy function $s(\rho) = \lim_{|N| \rightarrow \infty} \frac{1}{|N|} \log \mathcal{N}(\rho)$. Similarly, one can study the entropy of collaborative equilibria as a function of other parameters of the problem (e.g., costs of contribution) or topological property (e.g., average degree). Moreover, simple adaptations of the message-passing algorithm; e.g., by iteratively fixing some variables [BP-decimation (28)] or by introducing self-consistent biases on the messages [BP-reinforcement (29)], can be used to converge towards specific pure-strategy equilibria. In this respect, varying ϵ allows for searching collaborative equilibria with a given average density of collaborators ρ .

On the (ensemble of) subgraphs $(G, \bar{\Gamma})$ identified in this manner, the validity of the technical condition in Proposition 2 has to be assessed. It is possible to do it in polynomial time on single graphs or with probabilistic arguments on ensembles.

Results. A clear picture of the properties of collaborative equilibria can be obtained considering the conceptually simple situation of regular subgraphs of random regular graphs. It corresponds to assume a uniform cost; i.e., $X_i = X \forall i$, hence $\gamma_i = \gamma = \lceil X \rceil, \forall i$. The condition in Proposition 2 applies for $X_i = X < 3$, where $(G, \bar{\Gamma})$ is a collection of loops and dimers. It also applies for $X \geq 3$ with probability close to one, as the probability that a random regular graph of degree $\gamma \geq 3$ can be split in two disjoint component by removing one node, is negligible. Examples of collaborative equilibria obtained by means of the BP-decimation algorithm on a random regular graph of degree $|N_i| = K = 4$ are reported in Fig. 1. The figure suggests that

the density of collaborators ρ increases with the cost X_i . This conclusion is also supported by the behavior of the entropy $s(\rho)$, shown in Fig. 2. The entropy attains a maximum at a value ρ_{typ} of the density of collaborators, which implies that, for large $|N|$, almost all subgraphs span a fraction $\rho \simeq \rho_{\text{typ}}$ of the nodes of the network. Fig. 2 shows that ρ_{typ} increases with X_i , suggesting that higher costs enforce higher densities of collaborators. This is an apparently counterintuitive result that has however a simple explanation in the context of the model. An agent collaborates only as long as there are enough neighbors around her that also collaborate, and that would punish her defecting if she defects: the higher the cost of collaboration, the more profitable it would be to defect, but hence also the more collaborating neighbors are needed to balance this incentive.

A remarkable feature of Fig. 2 is that while for $\gamma = 1$ and 2 equilibria exist for all densities ρ of collaborators (i.e., $s(\rho) > 0, \forall \rho \in (0, 1]$), for $\gamma = 3$ equilibria only exist for $\rho \geq \rho_c \approx 0.8$. This behavior reflects the fact that for $\gamma = 1$, collaborative equilibria are collections of “dimers” of collaborators who reciprocally control each others in pairs. For $\gamma = 2$, equilibria are collections of loops, whose typical size in a sparse random graph is $\mathcal{O}(\log |N|)$. For large graphs, collaborative equilibria with $\gamma_i = 1$ or 2 can be constructed adding and removing dimers or loops, adjusting the density ρ in a continuous manner in $(0, 1]$. For $\gamma = 3$, instead, collaboration requires the formation of a regular subgraph of degree 3, which is only possible if more than a fraction ρ_c of the nodes are involved. Hence collaborative equilibria do not exist for $\rho < \rho_c$, and when they emerge (for $\rho > \rho_c$), they are exponentially many and they span a large fraction of the network. The existence of a threshold has clear implications for the systemic stability of collaborative equilibria: while a local deviation for $\gamma < 3$ only entails local rearrangements, for $\gamma \geq 3$ it is likely to cause the collapse of the whole collaborative structure. Hence collaborative equilibria acquire a nonlocal character whereby the fate of agreements in a neighborhood depend on what happens in distant regions of the social network. Such systemic fragility has its roots in the organization of the space of equilibria. Collaborative equilibria are formally obtained as solutions of constraint satisfaction problems (CSP) that belong to the class of locked CSP, recently introduced in ref. 27. In these CSP, if one modifies one variable in a solution, the rearrangement required to find another solution propagates across the network possibly affecting a sizable part of it. Here, for $\gamma < 3$, the shift from a given equilibrium to another one involves the rearrangement of at most $\mathcal{O}(\log |N|)$ nodes; i.e., a negligible fraction of the system. On the contrary, for $\gamma \geq 3$ the minimal distance between two equilibria is proportional to the number $|N|$ of nodes.

Fig. 2. Entropy $s(\rho)$ as a function of density ρ of collaborators for $\gamma = \lceil X \rceil = 1, 2$, and 3 on regular random graph with $|N_i| = K = 4$. Results are derived analytically for $|N| \rightarrow \infty$ as e.g., in ref. 27.

⁸⁸The probabilities $P\{i \in C\}$ and $P\{i \in \Gamma_j\}$ should be interpreted in a frequentist sense, as the fraction of CE in which $i \in C$ and $i \in \Gamma_j$. Their interpretation as mixed strategies can be misleading.

Fig. 3. (Left) Density ρ_{typ} of collaborators in a typical equilibrium in Erdős-Rényi graphs as function of the average degree $\langle k \rangle$ for $x = 0.1$ (black circles) and 0.5 (red squares). The inset shows the corresponding typical entropies $s_{\text{max}} = s(\rho_{\text{typ}})$. (Right) Probability $P_C(k)$ (red squares) that a node of degree $|N_i| = k$ collaborates in a typical collaborative equilibrium (ρ_{typ}) and low marginal cost ($x = 0.05$) on an uncorrelated scale-free random graph of size $|N| = 5 \cdot 10^3$ with degree distribution $P(k) \propto k^{-2.5}$ for $2 \leq k \leq 70$ (blue circles).

We have extended our analysis to heterogeneous networks; i.e., Erdős-Rényi and scale free random graphs with fixed costs $X_i = X$ [see also (26)] and with costs of collaboration $X_i = x|N_i|$ that are proportional to the number of neighbors each player interacts with. For these, the verification of the technical condition in Proposition 2 requires a more demanding analysis. We shall work under the assumption that this condition is verified, which means that our estimate of the number of equilibria can only be considered as an upper bound^{††}.

All the instances which have been analyzed confirmed the following set of generic features: (i) on average the typical fraction $\rho_{\text{typ.}}$ of collaborators increases with costs; (ii) the absence of equilibria at small densities ρ for large costs; and (iii) their fragility with respect to (w.r.t.) small perturbations and nonlocal character, for sufficiently large costs. When $X_i = x|N_i|$, if the marginal cost x is small, equilibria are mainly formed by dimers and loops and can be found for any density ρ . When x is large, nontrivial collaborative equilibria only exist for sufficiently large density of collaborators ρ , reproducing the “critical mass” effect observed in regular random graphs. Interestingly, when the marginal cost exceeds a graph-dependent threshold x_c ($x_c \simeq 0.79$ for Erdős-Rényi random graphs with average degree equal to 4), the number of equilibria vanishes in the full range of ρ (i.e., $s(\rho) < 0$), suggesting that the only possible equilibria are the all-defect or the fully collaborative (for $x = 1$) ones.

In addition, we also found that (iv) increasing the average degree promotes collaboration on average, because denser graphs admit typical equilibria of larger density ρ_{typ} of collaborators, as shown in Fig. 3 (left) for Erdős-Rényi random graphs. The monotonic behavior of $s(\rho_{\text{typ}})$ with graph density (see the inset) is an evidence of the fact that in denser graphs there are much more possible ways of arranging a collaborating subgraph. Finally, we found that (v) within a collaborative equilibrium, the more neighbors a player has, the more she is likely to collaborate (see Fig. 3 right).

Social networks are far from being uncorrelated. Therefore it is important to consider the role played by degree correlations. To this end, we have generated assortative/disassortative networks starting from uncorrelated ones by means of a link-exchange Monte Carlo algorithm proposed in ref. 30. The corresponding curves $s(\rho)$ computed for a moderate contribution cost $x = 0.1$ are displayed in Fig. 4. Our results suggest*** that (vi) positive

degree—correlation favors collaboration in that, the number of collaborative equilibria $s(\rho)$ and the typical fraction of collaborators ρ_{typ} increase with degree correlation. Remarkably, in strongly disassortative networks, collaboration can be suppressed altogether for large ρ (see Fig. 4).

Discussion

Experiments suggest that, apart from a very small fraction of innate altruists, most individuals are self-interested, hence they rationally condition their own collaboration to that of their peers (31). However, in order to sustain *conditional collaboration* over time in a social group, credible punishment is necessary. In a network setting, we have that it is sufficient for agents to condition their behavior only on the behavior of their neighbors. Conditional cooperation, in order to be credible, generally implies retaliation of defection only for a subset of neighbors. In other words, punishment is player-specific. This condition implies that, even in symmetric situations (e.g., complete network with $X_i = X$ for all $i \in N$), the equilibrium can be asymmetric, with heterogeneity of behaviors arising as an emergent property. Collaborative equilibria define a subgraph of pairwise interactions that reminds endogenous network formation games, in which stable coalitions depend on the existence of reciprocal pairwise interactions (32, 33).

In brief, our main result is that credibility and locality of threats identify a class of Nash equilibrium refinements, that we call “collaborative equilibria,” that can be described in terms of a purely graph theoretical problem. This result shows that the contribution cost (incentive to defect) has a major effect on the structure of the equilibria. When it is small, agents exert a low control on the neighbors and collaboration can be easily sustained in a repeated game. Increasing the cost/incentive, the system develops strong long-range correlations and *collaboration may require a critical mass*. Here, the absence of equilibria with low density of collaborators evokes Oliver and Marwell’s “critical mass theory,” which is one of the most celebrated theories of collective actions (34). The theory suggests that simultaneous coordination of a sufficiently large subset of individuals could be one of the possible solutions of the free-riding problem in situations in which individual contribution is extremely disadvantageous. Here we show that the existence of a critical mass emerges naturally from simple interaction, without the need of artificially introducing a threshold in the behavior of the agents. In this case, collaboration is very fragile and, even if there are exponentially many equilibria, learning to coordinate on one of them can be very difficult.

An important conclusion from our analysis, that partially contradicts results from evolutionary game theory and learning

[†]Again we expect that for small and for large X_i the condition should hold with high probability. Scale free random graphs with $X_i = x|N_i|$ is the most delicate case, where removing highly connected nodes might disconnect parts of the subgraph $(C, \tilde{\Gamma})$.

***Again, these results have to be taken with caution as they are derived assuming the validity of the technical assumption in Proposition 2.

Fig. 4. Entropy $s(\rho)$ as a function of density ρ of collaborators on correlated scale-free random graphs. We first considered an uncorrelated network ($J = 0$ in Eq. 9 of ref. 30) of $|N| = 5 \cdot 10^3$ nodes, with degree distribution $P(k) \propto k^{-2.5}$ for $2 \leq k \leq 70$ (black line and solid black diamond). Then we generated assortative/disassortative networks from this network, by means of a link-exchange Monte Carlo algorithm proposed in ref. 30, with positive ($J = 0.1$, triangles) and negative ($J = -1$, squares) degree-correlation. The cost is $X_i = x/|N_i|$ with $x = 0.1$.

dynamics based on imitation (9), is that network's sparseness and degree fluctuations do not necessarily favor the emergence of collaboration. In fact, the set of constraints imposed by control and punishment relations can be much more difficult to satisfy in sparse graphs than in dense groups. This result is somehow in agreement with experimental observation (10, 12), that measured systematically higher contribution levels in cliques and dense groups than in circular or linear arrangements.

The analysis of collaborative equilibria on different topologies suggests that the global properties of typical collaborative equilibria do not change considerably on different sparse networks. However, the local properties inside a network can be very different. It is evident in heterogeneous networks, where *high-degree*

nodes have a considerably larger probability to collaborate than low degree ones. The result is in agreement with predictions of evolutionary game theory (9) and with experimental results obtained comparing the contribution levels of central and peripheral nodes in star-like graphs (12). It also reminds the idea, firstly proposed by Haag and Lagunoff (35), of the existence of an uncollaborative fringe of agents connected to a collaborative core that can tolerate them. This picture is particularly true when considering heterogeneous networks with assortative mixing. On the contrary, in networks with negative degree correlations collaboration can be problematic, because of the mismatch between the conditions on high-degree nodes and the neighboring low-degree ones.

We conclude with a note on a recent experiment (36) on the absence of social contagion with respect to collaborative behavior. Social contagion refers to the idea (19) that some personal behaviors (as the inclination to collaborate) may be transmitted via social networks. Suri and Watts (36) find that increasing the number of collaborating neighbors does not directly imply a larger probability to collaborate (and vice versa). Our model provides a theoretical foundation for such an observation. In fact, strategic conditional collaboration requires a precise number of collaborating neighbors and when the level of collaboration in the neighborhood is too high the temptation to free-ride takes over. On the other hand, for sufficiently large costs, a collaborative state can be destroyed because of a single deviation, that may induce a cascade effect over the network.

In summary, we have put forward a unique framework to study the emergence of collaboration on networks, that is completely based on assumptions drawn from the experimental observations and on a rational forward-looking strategic behavior. Our results provide a theoretical explanation to a series of important empirical facts, and could motivate other experiments on repeated voluntary contribution games in networked systems. They provide insights in the analysis of strictly strategic problems, such as coalition formation, contractual agreements, and negotiation.

ACKNOWLEDGMENTS. L.D. acknowledges the support from the European Community grant STAMINA 265496.

- Jackson MO (2008) *Social and Economic Networks* (Princeton, New York).
- Homans CG (1958) Social Behavior as Exchange. *Am J Soc*, pp:597–606.
- Axelrod R (1984) *The Evolution of collaboration* (Basic Books, New York).
- Sobel J (2002) Can we trust social capital? *J Econ Lit* 40:139–154.
- von Neumann J, Morgenstern O (1944) *Theory of Game Econ Behav* (John Wiley and Sons, New York).
- Ellison G (1994) Collaboration in the prisoner's dilemma with anonymous random matching. *Rev Econ Stud* 61:567–588.
- Eshel I, Samuelson L, Shaked A (1998) Altruists, egoists, and hooligans in a local interaction model. *Am Econ Rev* 88:157–179.
- Nowak MA, May RM (1992) Evolutionary games and spatial chaos. *Nature* 359:826–829.
- Santos FC, Pacheco JM, Lenaerts T (2006) Evolutionary dynamics of social dilemmas in structured heterogeneous populations. *Proc Natl Acad Sci USA* 103:3490–3494.
- Kirchkamp O, Nagel R (2007) Naive learning and collaboration in network experiments. *Game Econ Behav* 58:269–292.
- Cassar A (2007) Coordination and collaboration in local, random and small world networks: experimental evidence. *Game Econ Behav* 58:209–230.
- Fatas E, Meléndez-Jiménez MA, Solaz H (2010) An experimental analysis of team production in networks. *Exp Econ* 13:399–411.
- Coleman JS (1988) Social capital in the creation of human capital. *Am J Sociol* 94: S95–S120.
- Burt RS (1992) *Structural Holes: The Social Structure of Competition* (Harvard University Press, Cambridge).
- Fudenberg D, Maskin E (1986) The folk theorem in repeated games with discounting or with incomplete information. *Econometrica* 54:533–554.
- Fehr E, Gächter S (2002) Altruistic punishment in humans. *Nature* 415:137–140.
- Mézard M, Parisi G, Zecchina R (2002) Analytic and algorithmic solution of random satisfiability problems. *Science* 297:812–815.
- Mézard M, Montanari A (2009) *Information, Physics, and Computation* (Oxford University Press, Oxford).
- Fowler JH, Christakis NA (2010) Collaborative behavior cascades in human social networks. *Proc Nat Acad Sci USA* 107:5334–5338.
- Rubinstein A (1986) Finite automata play the repeated prisoner's dilemma. *J Econ Theory* 39:83–96.
- Glance NS, Huberman BA (1993) The outbreak of collaboration. *J Math Sociol* 17:281–302.
- Benoit J, Krishna V (1993) Renegotiation in finitely repeated games. *Econometrica* 61:303–323.
- Kurzban R, Houser D (2001) Individual differences in collaboration in circular public goods game. *Eur J Personality* 15:37–52.
- Garey MR, Johnson DS (1979) *Computers and Intractability* (Freeman, New York).
- Marinari E, Semerjian G (2006) On the number of circuits in random graphs. *J Stat Mech* P06019.
- Pretti M, Weigt M (2006) Sudden emergence of q-regular subgraphs in random graphs. *Europhys Lett* 75:8–14.
- Zdeborová L, Mézard M (2008) Constraint satisfaction problems with isolated solutions are hard. *J Stat Mech* P12004.
- Braunstein A, Mézard M, Zecchina R (2005) Survey propagation: an algorithm for satisfiability. *Random Struct Algor* 27:201–226.
- Braunstein A, Zecchina R (2006) Learning by message passing in networks of discrete synapses. *Phys Rev Lett* 96:030201.
- Noh JD (2007) Percolation transition in networks with degree-degree correlation. *Phys Rev E* 76:026116.
- Fischbacher U, Gächter S, Fehr E (2001) Are people conditionally collaborative? Evidence from a public goods experiment. *Econ Lett* 71:397–404.
- Jackson MO, Wolinsky A (1996) A Strategic model of social and economic networks. *J Econ Theory* 71:44–74.
- Vega-Redondo F (2006) Building up social capital in a changing world. *J Econ Dyn Control* 30:2305–2338.
- Marwell G, Oliver P (1993) *The Critical Mass in Collective Action: A Micro-Social Theory* (Cambridge University Press).
- Haag M, Lagunoff R (2006) Social norms, local interaction and neighborhood planning. *Int Econ Rev* 47:265–296.
- Suri S, Watts D (2011) Cooperation and contagion in web-based, networked public goods experiments. *PLoS ONE* 6:e16836.