Can Reinforcement Learning Enhance Social Capital?

He Zhao[1], Hongyi Su[1], Yang Chen[2][0000−0002−1148−3920], Jiamou Liu[2][0000−0002−0824−0899], Bo Yan[1], and Hong Zheng[1]

[1] Beijing Lab of Intelligent Information Technology, School of Computer Science, Beijing Institute of Technology, Beijing, China
{2120171104,henrysu,yanbo,hongzheng}@bit.edu.cn
[2] School of Computer Science, The University of Auckland, Auckland, New Zealand
{yang.chen,jiamou.liu}@auckland.ac.nz

Abstract. Social capital captures the positional advantage gained by an individual by being in a social network. A well-known dichotomy defines two types of social capital: *bonding capital*, which refers to welfare such as trust and norms, and *bridging capital*, which refers to benefits in terms of influence and power. We present a framework where these notions are mathematically conceptualized. Through the framework, we discuss the process when an individual gains social capital through building new edges. We explore two questions: (1) How would an individual optimally form new relations? (2) What are the impacts of the network structure on the individual's social capital? For these questions, we adopt a paradigm where the individual is a utility-driven agent who acquires knowledge about the network through repeated trial-and-error. In this paradigm, we propose two reinforcement learning algorithms: one guarantees the convergence to optimal values in theory, while the other is efficient in practice. We conduct experiments over both synthetic and real-world networks. Experimental results indicate that a centralized structure can enhance the performance of learning.

Keywords: social capital · reinforcement learning · network building · social network.

1 Introduction

Human societies are the products of interactions among social actors. The reward theory of attraction, well-established in social psychology, states that people tend to interact with those whose behavior is rewarding to themselves or those who are associated with rewarding events [17]. To study social dynamics, it is thus important to understand the essential constituents of rewards that motivate social networking, and how these rewards impact an individual's decision making. Social networking not only brings an individual tangible rewards such as economic resources, but also intangible benefits in the form of trust, social support, information control, and social influence. The notion of *social capital* was introduced to unify these intangible benefits [9]. While undoubtedly a significant portion of social interactions are driven by concerns of social capital, the direct link between social capital and social interaction is often overlooked. In this paper, we tackle this issue through investigating the *network building* process, where an individual in a social network purposely establish social interactions in the hope of gaining social capital.

Our investigation aims to resolve several challenges around the network building process. The first seeks to pinpoint the notion of social capital. A well-known dichotomy has divided discussions on social capital into two categories: *bonding capital* and *bridging capital* [15]. The former depicts the aggregate welfare that an individual draws from its closed social circle in the form of, e.g., trust and social support [5], while the latter captures the individual's capacity to acquire opportunities and information via open links and determines, e.g., status and power [7]. An individual's reward in social networking would be a combination of these two forms of social capital.

The second concerns optimizing the individuals' social networking tactics when establishing social interactions. By this, We would like to identify the optimal decisions subject to the following constraints: (1) Since rewards are typical of hindsight, an individual demands learning from experience via trial-and-error. (2) An individual has limited abilities to establish and maintain relationships. Since if otherwise, trivially she can link to all others to maximize rewards. (3) The individual would have only partial observation about the network and consequently social interaction can only be established with those that are in the surrounding social circle.

The third aims to develop insights regarding how the network structure impacts an individual's social capital. A network may exhibit some salient structural properties such as community structure [11], scale-free property [3] and small-world property [25]. Intuitively, an individual may relatively easily gain social capital in networks with one or multiple centers, as she can always fast approach to these centers or act as bridges between centers. In contrast, a homogeneous network imposes difficulties over making decisions for establishing links, which implies that gaining social capital is hard in such networks.

Contribution. Motivated by the concerns above, we propose the *social capital-driven network building* (SCNB) problem, which involves a social network and an individual within it. The individual has partial observation of the network and can create a limited number of edges, denoting social interactions. As the agent has imperfect information about rewards and network structures, we tackle SCNB problem using the reinforcement learning paradigm. That is, the agent learns to gain social capital via trial-and-error and referring to previous experiences. We highlight the novelty of this work as follows: (1) We present a framework where bonding and bridging capitals are mathematically conceptualized. (2) We propose a fast Q-learning algorithms FQL for SCNB. FQL's efficiency makes it applicable to networks of a non-trivial size. Not only the algorithm leads to quick stability, it also has a performance also compares well with the optimal actions, as computed by a much slower alternative OQL. (3) We conduct experiments on synthetic and real-world networks. Results reveal that centralized network structures enable an agent to gain social capital more effectively.

Related Works. Pioneering works of sociologists advanced the research on social capital. Coleman's serial works lay the foundation for the research on social capital [9, 8]. Bourdieu proposes that homophily is the source of bonding capital [5]. Granovetter, Putnam, and Burt state that weak ties are the source of bridging capital [12, 19, 7]. Recently, a vast literature uncovers vital roles that social capital plays in a wide range of fields such as resource management [4], disaster survival [2] and youth unemployment [13].

Game-based research on *network formation* focuses on equilibria among rational agents [14, 6, 21], where behaviors of agents are subject to restricted predefined rules. Our work surpasses theirs as the learning process captures initiatives of agents. Particularly, authors in [1] propose a micro-founded mathematical model of the evolution driven by social capital, where bridging capital is defined as the betweenness centrality. In this paper, we formalize bridging capital following their work.

Algorithmic research on *network building* problem asks for integrating a newcomer to the center of an existing static or dynamic network via establishing a minimum number of links [16, 26, 27]. Network building problem is computationally hard and thus several efficient heuristic-based algorithms are given in these works. Our work differs from theirs as: (1) The objective of the agent is to gain social capital rather than centrality; (2) We assume that an agent does not have global knowledge of the network, which adheres more to reality; (3) We solve SCNB using reinforcement learning instead of heuristic-based algorithms.

2 Preliminaries

2.1 Social Capital

Rooted in sociology, the concept of social capital aims to capture the benefits attained by individuals via social interactions. Such benefits can emerge in the form of social support, companionship, solidarity, influence, and control over information, which are closely related to network structural properties. As a result, social capital should be measured based on structural properties. However, the structure itself does not define social capital. Instead, social capital arises as a function of social interactions and information conveyed through social relations [8]. Social relations have long been classified based on their functions: While strong ties link homogeneous and like-minded individuals, weak ties bridge diverse and weakly connected groups [12]. Analogously, social capital also consists of two types: *Bonding capital* refers to benefits an individual draws from its closed neighborhood, in the form of, e.g., trust and support, which are brought by strong ties; while *bridging capital* is an incarnation of benefits of accessibility to information and control over information flow, which are largely functions of weak ties. However, so far no consensus has been reached over the formal definitions of social capital. In this paper, we adopt two metrics to measure these two notions in the context of social networks.

A social network captures repeated patterns of social interactions and is viewed as an undirected graph $G = (V, E)$, where V is a set of nodes (agents) and $E \subseteq V^2 \setminus \{uu \mid u \in V\}$ is a set of edges. A (k-length) *path* is a sequence of nodes $u_1u_2, \ldots u_{k+1}$ where $u_iu_{i+1} \in E$ for $1 \leq i \leq k$. The distance $\mathsf{dist}_G(u, v)$ between u and v is the length of a shortest path between these two nodes. We focus on connected graphs and thus $\mathsf{dist}_G(u, v) < \infty$ for any $u, v \in V$. If $\mathsf{dist}_G(u, v) = d$, then we say that u, v are *d-hop neighbors*. The *d-hop neighbor set* of $v \in V$ is $N_G^d(v) \coloneqq \{u \in V \mid \mathsf{dist}_G(u, v) = d\}$.

In the real-world, the sight of an individual tends to be restricted by the large scale and complex dynamics of social networks, resulting in incomplete knowledge of the environment. Consequently, an agent often has a good understanding of its neighborhood but is unfamiliar with distant parts of the network. To capture this fact, we employee the notion of *2-level ego network*, which represents the social surrounding that an agent perceives and maintains.

Definition 1 (2-level Ego Network). *The* 2-level ego network *of a node $v \in V$ is the subgraph O_G^v of G induced by v, v's 1-hop neighbors, and 2-hop neighbors, i.e., the nodes of O_G^v is $V_G^v = \{u \in V \mid \mathsf{dist}_G(u, v) \leq 2\}$ and the edges of O_G^v is $E_G^v = \{uw \in E \mid u, w \in V_G^v\}$. The node v is called the* ego *in O_G^v.*

Considering distinct differences between origins of bonding and bridging capital, we formalize them using two standard metrics, *personalized PageRank* and *betweenness centrality*, respectively. Bonding capital, as it is often expressed as trust and companionship, measures the degree to which two nodes bind with each other [5]. This can be aptly captured through a measure of "social proximity". In other words, a node gains more bonding capital as it gets closer to others in its own neighborhood. Personalized PageRank index is adapted from PageRank and evaluates structural proximity between nodes through predicting the likelihood of edges between any pairs of nodes [18]. It takes as input a starting node s, and assigns a score to every node u that captures the likelihood of a random walk from s to reach u [23]. Fix a restart probability $\beta \in (0, 1)$, random walk starts from the node s[3]; stops moving at each node with the probability of β and restarts from the node s; or continues to walk with the probability of $1 - \beta$ by randomly selecting a node from the neighbors of the current node. The probability that each node is accessed converges in finite rounds of walking. Each entry of the personalized PageRank vector $\boldsymbol{pr}$ records the probability that the corresponding node is accessed. More formally, let $\boldsymbol{a}_u$ be the column vector in the adjacency matrix of G corresponding to node u. Denote by pr_u, the link prediction score between s and u as:

$$\mathsf{pr}_u = \beta r_u + (1 - \beta)(\boldsymbol{pr} \cdot \boldsymbol{a}_u / |N_G^1(u)|), \tag{1}$$

where $r_u = 1$ if $u = s$ and $r_u = 0$ otherwise. Intuitively, assume that s holds certain amount of "goodwill" which is randomly shared with s's neighbors, and whoever that obtains such goodwill can continue to pass goodwill to their neighbors or return them to the node s, in the same manner as a random walk as

[3] In experiments, we set the restart probability $\beta = 0.15$.

above. Bonding capital can be viewed as the amount of goodwill eventually received by s.

Definition 2 (Bonding Capital). *Given a graph $G = (V, E)$ and a node $v \in V$, denote by $\mathsf{bo}_G(v)$, the* bonding capital *of v is defined by summing personalized PageRank indices between v and v's neighbors, namely,* $\mathsf{bo}_G(v) := \sum_{u \in N_G^1(v)} \mathsf{pr}_u$.

Occupying a central position to act as a gateway for information exchange brings an individual bridging capital [7]. Furthermore, betweenness centrality is used to evaluate bridging capital in [1]. The betweenness centrality of a node measures the number of the shortest paths between each pair of other nodes that pass through it, and thus reflects the agent's ability to broker interactions between different groups of agents. Following the work [1], we formalize bridging capital using betweenness centrality.

Definition 3 (Bridging Capital). *Let $G = (V, E)$ be a connected graph. The* bridging capital *of $v \in V$, denote by $\mathsf{br}_G(v)$, is defined as v's betweenness centrality: $\mathsf{br}_G(v) := \sum_{s \neq v \neq t \in V} \sigma_{st}(v)/\sigma_{st}$, where σ_{st} is the number of shortest paths between nodes s and t, and $\sigma_{st}(v)$ is the number of shortest paths passing v.*

An individual may show different preferences to two types of capital. To cope with this, we employee a *preference weight* $w \in [0, 1]$ to define the *mixed capital*.

Definition 4 (Mixed Capital). *Let $G = (V, E)$ be a network. For a node $v \in V$ and a preference weight $w \in [0, 1]$, the* mixed capital *induced by w is defined as* $\mathsf{mix}_G^w(v) := w\mathsf{bo}_G(v) + (1 - w)\mathsf{br}_G(v)$.

2.2 Q-Learning

We next briefly state backgrounds of Q-learning that defines a learning method for Markov decision processes (MDPs). See [22] for a thorough introduction. An MDP is a tuple $\langle \mathcal{S}, \mathcal{A}, r, p \rangle$, where $\mathcal{S}$ is the discrete state space, $\mathcal{A}$ is the discrete action space, $r : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function, and $p : \mathcal{S} \times \mathcal{A} \rightsquigarrow \mathcal{S}$ is the transition function, where $\rightsquigarrow$ denotes a probabilistic mapping.

To solve an MDP, we learn estimates for the optimal value of each action, defined as the expected sum of long-term rewards when taking the action and following the optimal policy thereafter. Under a given *policy* $\pi : \mathcal{S} \to \mathcal{A}$, the true value of an action a in a state s is

$$Q^\pi(s, a) := \mathbb{E}\left[R_1 + \gamma R_2 + \gamma^2 R_3 + \cdots \middle| S_0 = s, A_0 = a, \pi\right], \tag{2}$$

where $\gamma \in [0, 1]$ is the *discount factor* and S_t, A_t, R_t denote the state, action and reward at time t, respectively. The optimal value is then $Q^*(s, a) = \max_\pi Q^\pi(s, a)$. An optimal policy π^* is derived from optimal values by choosing highest-valued actions in each state. Estimates for optimal action values can be learned using Q-learning. At each time t, the agent maintains a state-action value Q_t. In a state s, the agent takes an action a, receives a reward r', and enters a next state s'. Then, an action-value is updated using the following rule:

$$Q_{t+1}(s,a) = (1-\alpha_t)Q_t(s,a) + \alpha_t \left[r' + \gamma \max_{a'} Q_t(s',a') \right], \quad (3)$$

where $\alpha_t \in [0,1)$ is the *learning rate*. It has been proved that the sequence $(Q_t(s,a))_{t\in\mathbb{N}}$ generated by Eq. (3) converges to the optimal $Q^*(s,a)$ [24] with following assumptions on the learning rate:

$$\sum_{t=1}^{\infty} \alpha_t(a) = \infty \qquad \text{AND} \qquad \sum_{t=1}^{\infty} \alpha_t^2(a) < \infty. \quad (4)$$

3 Problem Setup

A question naturally arises as to how would an individual gain social capital in a partially perceived society. Throughout the remainder of this paper, we use v to denote a *learner* in a network $G = (V, E)$, who aims to gain social capital through establishing new social interactions, which take the form of extra edges to G. Formally, for $uv \notin E$, $G \oplus_v u$ denotes the network $(V, E \cup \{uv\})$. We formalize the interaction between v and the network G as a process happening in (finite) discrete time steps $\tau = 0, 1, 2, \ldots$. At each time step, v establishes an edge with an agent selected from its 2-level ego network, resulting in an updated network. This process is formally defined below.

Definition 5 (Network Building Process). *An (ℓ-round)* network building (NB) process *between $G = (V, E)$ and $v \in V$ consists of a finite sequence of networks $G_0, G_1 \ldots, G_\ell$ and a finite sequence of nodes $u_0, u_1, \ldots, u_{\ell-1}$ such that $\forall\, 0 \le \tau \le \ell - 1$: $u_\tau \in N^2_{G_\tau}(v)$, where $G_0 = G$ and $G_{\tau+1} = G_\tau \oplus_v u$.*

Definition 6 (Network Building Strategy). *A* network building strategy *is a function φ that takes as input a 2-level ego network O^ν_G and outputs a node $u \in N^2_G(\nu)$. Any NB process $(G_0, G_1 \ldots, G_\ell, u_0, u_1, \ldots, u_{\ell-1})$ is said to be* consistent *with a strategy φ if $\forall 0 \le \tau \le \ell - 1$: $u_\tau = \varphi(O^v_{G_\tau})$.*

We write $\mathcal{G}_\tau$ for the space of all networks that might emerge at time step τ and $\mathcal{G} = \bigcup_{1\le\tau<\ell} \mathcal{G}_\tau$ for all possible realizations of networks during the NB process. The space of v's all possible 2-level ego networks is therefore $\mathcal{O} := \{O^\nu_H \mid H \in \mathcal{G}\}$. Now we are ready to formally define the SCNB problem.

Definition 7 (Social Capital-based Network Building Problem). *A* social capital-based network building *(SCNB) problem is a tuple $\langle G, \nu, \ell, w\rangle$, where $G = (V, E)$ is a graph, $\nu \in V$ is the learner, $\ell \in \mathbb{N}^+$ and $w \in [0,1]$. The problem asks for an NB strategy φ^* so as to maximize the mixed capital of $\mathsf{mix}^w_{G_\ell}(\nu)$ via an ℓ-round NB process consistent with φ^*.*

The SCNB problem has the inherent high computational complexity. The *network building* problem – asking for a minimum vertex set of a static network to minimize a newcomer's eccentricity via linking to all nodes in the set – has been shown to be NP-complete in [16] where several heuristic-based algorithms are given. SCNB is more complicated as the partial observation imposes an extra layer of complexity. Hence the heuristic-based algorithms are no longer effective as they require global information in advance. Moreover, to cope with the incomplete knowledge of the network, an agent needs to learn from the environment by trials and retrieving previous experiences. Motivated by above discussions, it is the most naturally to adopt reinforcement learning to solve SCNB problem, in which an agent learns a strategy via trial-and-error.

4 Reinforcement Learning for SCNB

This section studies reinforcement learning algorithms for SCNB. Let $\langle G, \nu, \ell, w\rangle$ be a SCNB problem. Learner v iteratively refines a network building strategy φ through replay and trail-and-error. A trial of a complete ℓ-round network building process is called an *episode*. We first show the formulation of reinforcement learning in the context of SCNB problem. Then we propose two Q-learning algorithms for SCNB.

4.1 Reinforcement learning formulation

We define the states, actions, rewards, transitions and policies in reinforcement learning for SCNB as follows:

1. **States:** In each episode, for any round $0 \leq \tau < \ell$, v utilizes an *ego-state mapping* ϕ to extract a state S_τ from the current 2-level ego network $O^v_{G_\tau}$.

   Definition 8 (Ego-State Mapping). *An* ego-state mapping *is a function $\phi : \mathcal{O} \to \mathcal{S}$ that maps a 2-level ego network to a state in a finite state set $\mathcal{S}$.*

2. **Actions:** An action A_τ is a node from v's 2-hop neighbor set $N^2_{G_\tau}(v)$.
3. **Rewards:** A reward function $r(S_\tau, A_\tau)$ at state S_τ is defined as the change of the mixed capital as a function of taking action A_τ and entering into next round $\tau + 1$, that is,

$$r_{\tau+1} = r(S_\tau, A_\tau) := \mathsf{mix}^w_{G_{\tau+1}}(\nu) - \mathsf{mix}^w_{G_\tau}(\nu). \tag{5}$$

4. **Transition:** The transition here is deterministic. Given the current state S_τ and action A_τ, the process transits into next state with probability 1, i.e.,

$$p(S_\tau, A_\tau) = S_{\tau+1} = \phi(O^v_{G_\tau}). \tag{6}$$

5. **Policy:** A policy π takes input as a state S_τ and outputs a action A_τ.

Conceptually, based on state s_τ, v chooses a node a_τ (action) from v's 2-hop neighbor set $N^2_{G_\tau}(v)$ to create a link. Then, v receives a reward $r_{\tau+1} = \mathsf{mix}^w_{G_{\tau+1}}(v) - \mathsf{mix}^w_{G_\tau}(v)$ (change of mixed capital) and enters into the next round $\tau + 1$. We use $Q(S_\tau, A_\tau)$ to denote the action-value of the state-action pair (S_τ, A_τ). Note that since ℓ is finite, trivially we define $Q(S_i, A_i) = 0, \forall i \geq \ell$. For exploration, we employee the standard ε-greedy method as the policy π derived from Q. Namely, choose actions greedily most of the time, but with a very small *exploration probability* ε choose randomly. Formally,

$$A_\tau = \pi(S_\tau) = \begin{cases} \arg\max_{a \in N^2_{G_\tau}(v)} Q(S_\tau, a) & \text{with probability } 1-\varepsilon \\ \text{randomly selected } a \in N^2_{G_\tau}(v) & \text{with probability } \varepsilon \end{cases} . \tag{7}$$

Throughout, we set the discount factor $\gamma = 1$, i.e., the future rewards are not discounted. Thus the long-term cumulative reward in Q-learning is equivalent to the mixed capital after the last round in PONB. Therefore naturally we have the following lemma.

Lemma 1. *The optimal policy π^* in Q-learning for PONB is equivalent to an optimal NB strategy φ^* if the discount factor $\gamma = 1$.*

A strategy φ is obtained after several rounds of learning such that $\forall 0 \leq \tau < \ell : \varphi(O^v_{G_\tau}) = \pi(S_t) = \pi(\phi(O^v_{G_\tau}))$. We next show two Q-learning algorithms for SCNB that are separated by two different ego-state mappings.

4.2 First Try: An optimal Q-learning algorithm for SCNB

We first investigate the optimal Q-learning algorithm (OQL) that directly uses 2-level ego networks as states. More formally, for a SCNB problem $\langle G, \nu, \ell, w\rangle$, OQL utilizes an ego-state mapping ϕ_o such that $\forall 0 \leq \tau < \ell : \phi_o(O^\nu_{G_\tau}) := O^\nu_{G_\tau}$. According to Eq. (3) and Lem. 1, the Q-value update rule of OQL is

$$Q_{t+1}\left(O^v_{G_\tau}, a\right) = (1-\alpha_t) Q_t\left(O^v_{G_\tau}, a\right) + \alpha_t \left[r_{\tau+1} + \max_{a'} Q_t\left(O^v_{G_{\tau+1}}, a'\right)\right], \tag{8}$$

where $a = A_\tau \in N^2_{G_\tau}(\nu)$, $G_{\tau+1} = G_\tau \oplus_v a$ and $a' \in N^2_{G_{\tau+1}}(v)$. The following theorem shows the optimality of OQL.

Theorem 1. *The sequence $\left(Q(O^v_{G_\tau}, a)\right)_{t\in\mathbb{N}}$ generated by Eq. (8) converges to the optimal $Q^*(O^\nu_{G_\tau}, a)$ under the assumptions stated in Eq. (4).*

Proof. We complete this proof by showing that the underlying process of OQL is an MDP. Let $\langle G, v, \ell, w\rangle$ be a PONB problem. Construct an MDP as follows: The state set is the space of the 2-levle ego network $\mathcal{O}$. The action set is $\mathcal{A} := \{u \in V \mid \mathsf{dist}(u, v) \geq 2\}$, i.e., all nodes other than v and v's 1-hop neighbors. Note that at each time step τ, the set of eligible actions $N^2_{G_\tau}$ is a subset of $\mathcal{A}$. The definitions of reward function r and transition function p remain unchanged as stated above. It is notable that p degenerates to a deterministic mapping in this setting. The tuple $(\mathcal{O}, \mathcal{A}, r, p)$ is clearly an MDP and hence the optimality of OQL holds. □

After convergence of all Q values we obtain a near-optimal policy $\pi = \pi^*$ and therefore get a near-optimal NB strategy $\varphi = \varphi^*$ (according to Lem. 1). The procedure of OQL is shown in Alg. 1.

Algorithm 1: OQL: Estimate optimal strategies for SCNB

Input: An SCNB problem $\langle G, v, \ell, w\rangle$ with $G = (V, E)$
Output: An NB strategy $\varphi = \varphi^*$
Initialization: Set $Q(o, a) = 0$ for all $o \in \mathcal{O}, a \in V$
for *each episode* **do**
  $o \leftarrow O^v_G$, $F \leftarrow E$
  for *round* $i = 0 \to \ell - 1$ **do**
    $a \leftarrow \pi(o) = \varphi(o)$ ▷ ε-greedy
    $r \leftarrow \mathsf{mix}^w_{(V,F\cup\{va\})}(v) - \mathsf{mix}^w_{(V,F)}(v)$
    $F \leftarrow F \cup \{va\}$
    $o' \leftarrow O^v_{(V,F)}$
    $Q(o, a) \leftarrow (1 - \alpha)Q(o, a) + \alpha\,[r + \max_{a'} Q(o', a')]$
    $o \leftarrow o'$

4.3 A fast Q-learning algorithm for SCNB

However, OQL is impractical for most of the time as the space of 2-level ego networks is typically too large to learn from. Therefore we propose a fast Q-learning algorithm (FQL) that aggregates all 2-level ego networks at round τ to a single state. In other words, FQL directly uses the indices of time steps as states. Formally, FQL utilizes an ego-state mapping ϕ_f such that $\phi_f(O^v_{G_\tau}) := \tau$ for all $G_\tau \in \mathcal{G}_\tau$. Hence the Q-value update rule of FQL is

$$Q_{t+1}(\tau, a) = (1 - \alpha_t)Q_t(\tau, a) + \alpha_t \left[r_{\tau+1} + \max_{a'} Q_t(\tau + 1, a') \right], \tag{9}$$

where $a = A_\tau \in N^2_{G_\tau}(v)$, $G_{\tau+1} = G_\tau \oplus_v a$ and $a' \in N^2_{G_{\tau+1}}(v)$.

The reward distribution is not stationary in FQL as the underlying network $G_{\tau+1}$ induced by a same action a at time τ may vary. Hence $Q(\tau, a)$ is not guaranteed to converge to $Q^*(O^v_{G_\tau})$ for any $G_\tau \in \mathcal{G}_\tau$. In other words, the cost of a fast speed is losing the guarantee of optimality. However, surprisingly, we observe that FQL often stablizes faster compared to OQL, which implies that FQL achieves a good trade-off between time and accuracy in practice (see Sec. 5). As the guarantee of optimality is absent in FQL, we obtain a sub-optimal policy π and thus a sub-optimal NB strategy φ. Alg. 2 shows the procedure of FQL.

5 Simulations and Experiments

We simulate the SCNB processes and test two Q-learning algorithms over both synthetic and real-world networks. As a uniform setting, we set learning rate $\alpha = 0.1$ and exploration probability $\varepsilon = 0.3$, respectively. The discount factor γ is fixed to 1 as addressed in Sec.4. To capture how a marginal individual embeds herself to a society through gaining social capital, throughput experiments, we determine v as the node with the lowest degree in the corresponding network.

Algorithm 2: FQL: Estimate sub-optimal strategies for SCNB

Input: An SCNB problem $\langle G, v, \ell, w\rangle$ with $G = (V, E)$
Output: A sub-optimal NB strategy φ
Initialization: Set $Q(i, a) = 0$ for all $0 \leq i \leq \ell, a \in V$
for *each episode* **do**
  $F \leftarrow E$
  for *round* $i = 0 \to \ell - 1$ **do**
    $a \leftarrow \pi(i) = \varphi(i)$ ▷ ε-greedy
    $r \leftarrow \mathsf{mix}^{w}_{(V,F\cup\{va\})}(v) - \mathsf{mix}^{w}_{(V,F)}(v)$
    $F \leftarrow F \cup \{va\}$
    $Q(i, a) \leftarrow (1 - \alpha)Q(i, a) + \alpha\,[r + \max_{a'} Q(i + 1, a')]$

5.1 Football teams network: a case study

The goal is to test OQL and FQL on a real-world network. We apply two Q-learning algorithms to the FOOTBALL dataset, which represents American football games between Division IA colleges during regular season Fall 2000. Nodes and edges represent teams and matches, respectively [20]. In this scenario, we assume a football team of a college is seeking for a match (creating a link) with another college. The pursuit of bonding capital of a team can be interpreted as exploiting the league that it belongs to. By contrast, the bridging capital can be seen as the capacity to broker two different leagues.

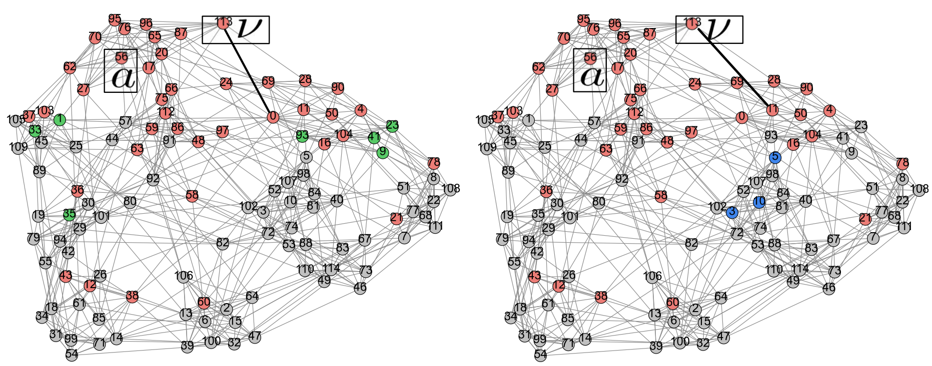

Fig. 1. The agent v and action a are set as nodes 113 and 56, respectively. Red nodes represent v's initial 2-hop neighbors. Green nodes in the left figure represent v's new 2-hop neighbors after linking to node 0 at round 1. Blue nodes in the right figure represent 2-hop neighbors after linking to node 11 at round 1. o_1 is induced by red and green nodes (left figure), and o_2 is induced by red and blue nodes (right figure).

This simulation is configured as follows. We set the length of NB processes $\ell = 5$ and the preference weight $w = 0.5$. We select two possibly emerging 2-level ego networks (denoted by o_1 and o_2) at round $\tau = 1$, and a node a (see

Fig. 1). We execute 10 independent runs for this experiment. The learning curves of $Q(o_1, a)$ and $Q(o_2, a)$ in OQL, $Q(\tau = 1, a)$ in FQL, and tendencies of social capital are plotted in Fig. 2.

Two facts stand out from experimental results: (1) FQL standouts as $Q(1, a)$ stabilizes considerably faster than two Q-values in OQL, though to a non-optimal value. This coincides with our theoretical results. (2) Thanks to the fast stabilization, FQL surpasses OQL in the speed of enhancing social capital. More surprisingly, FQL stabilizes at a nearly optimal value of mixed capital, as shown in the right part of Fig. 2. Therefore, it is explicitly that FQL successfully achieves a trade-off between efficiency and accuracy in this case study.

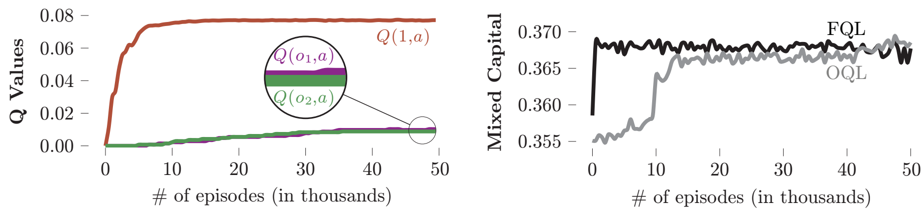

Fig. 2. Results of applying two Q-learing algorithms on football network. **Left:** Q-values by OQL ($Q(o_1, a)$ and $Q(o_2, a)$) and FQL ($Q(1, a)$). Each curve represents the median over 10 independent runs. **Right:** Results of mixed capital by OQL and FQL with preference weight $w = 0.5$.

5.2 Random networks

The objective of this experiment is to investigate how social structure influences the learning speed and effect. We only use FQL in this experiment as OQL requires impractical running time. We conduct simulations on synthetic networks that are generated by standard random network formation model: **(1) Barabási-Albert (BA) model.** Barabási-Albert model generates scale-free networks with a power-law degree distribution, which exhibit in general low clustering coefficient. The process initiates a cycle of n nodes, and from each node, it adds at most $m < n$ edges to others using a preferential attachment scheme. Following the convention, we set $m = 4$ in our experiments [3]. **(2) Watts-Strogatz (WS) model.** The model starts from a regular ring lattice and randomly rewires edges with probability p. As edges are rewired, the network swings from a regular graph towards an ER random graph. When $p \in [0.01, 0.1]$, the graph typically exhibits a high clustering coefficient, demonstrating a small world property [25]. To better reflect the small world property, we set $p = 0.1$ in our experiments. **(3) Erdös-Rény (ER) model.** The ER model builds edges as independent Bernoulli random variables and results in Poisson binomial degree distribution, with low clustering coefficient in general [10]. The model starts with a number n of agents and adds edges between pairs of agents uniformly randomly with probability $p \in [0, 1]$ [10]. Here we set $p = 0.04$. We generate an instance for each model, each of size 500. Tab. 1 shows the statistics of three random networks.

Table 1. Properties of three random networks.

network model	nodes	edges	radius	diameter	clustering coefficient
BA	500	1491	4	6	0.05
WS	500	2000	6	8	0.49
ER	500	4964	3	4	0.04

The length of NB processes ℓ is set as 5 and the preference weight w is varied in $\{0, 0.5, 1\}$. The values of social capital shown in Fig. 3 are averaged over 10 independent runs. We make two discussions: • A centralized network structure can increase learning efficiency. Stabilization in ER networks requires a large number of episodes (more than 7000). While both types of social capital in BA and WS stabilizes very fast (at ≈ 2500 episodes). • A centralized network structure can enhance the effect of learning. Bonding and bridging capital in BA and WS networks stabilize faster than in ER networks. Moreover, in ER networks, bonding and bridging capital rise by merely ≈ 0.004 and 0.04 after 10000 episodes, respectively, which is considerably lower than in BA and WS networks. This can be explained from the aspect of network structure: BA networks have one and multiple centers; nodes in WS networks are connected densely; while nodes in ER networks are randomly and loosely connected. This captures the fact that an individual can fast learn to access to some center nodes in a densely connected society but fails to remarkably enhance social capital in a chaotic and loosely connected society.

5.3 Real-world networks

Following the same setting of experiments on synthetic networks, we next simulate NB processes on three real-world networks[4]: **(1) FRIEND.** Friend network contains friendships between users of a social website, where nodes and edges represent users and friendships, respectively. **(2) EMAIL.** This network is generated using email data from a large European research institution, which studies anonymized information about all incoming and outgoing emails between institutional members. There is an edge (u, v) in the network if person u sent person v at least one email. **(3) BIBLE.** Bible network contains nouns (places and names) of the King James Version of the Bible and information about their co-occurrences. A node represents one of the above noun types, and an edge indicates that two nouns appeared together in the same Bible verse. Tab. 2 summaries the statistics of three above real-world networks.

[4] All three real-world network datasets are from the public Koblenz Network Collection. `http://konect.uni-koblenz.de/networks/`

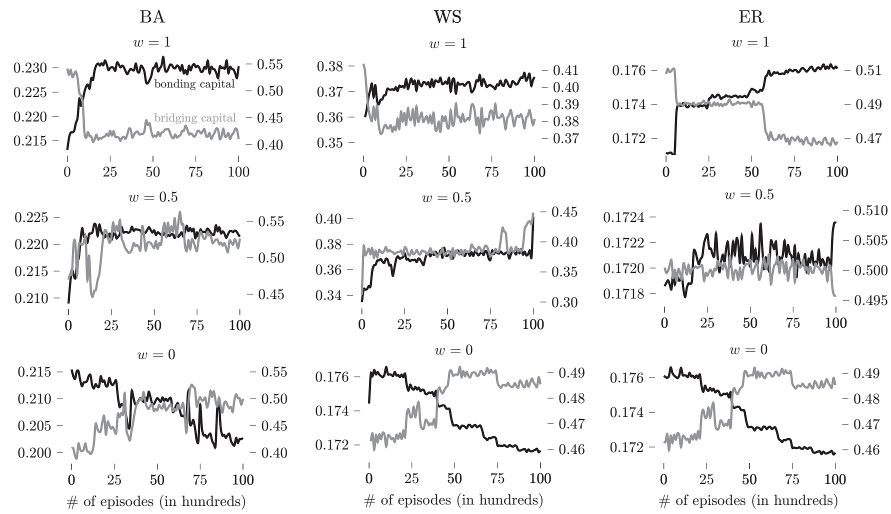

Fig. 3. Results of bonding (left axis and dark lines) and bridging capital (right axis and gray lines) in BA, WS and ER networks during 10000 episodes with varying preference weight w. Each curve is averaged over 10 independent runs.

The results on three real-world networks are illustrated in Fig. 4. Analogous to the results in random networks, the performance of the learning exhibits a positive relationship with the clustering coefficient, a measure of the extent to which nodes in a graph tend to cluster together. The FRIENDS network has the lowest clustering coefficient, where social capital stabilizes after ≈ 8000 episodes. In contrast, the learning speed in the other two networks is relatively higher, where stabilization can be observed before ≈ 5000 episodes (EMAIL with $w = 1, w = 0.5$; BIBLE with $w = 1, w = 0.5$). This result is consistent with our discussions in the experiment on random networks that a centralized structure can enhance the performance of learning.

Table 2. Properties of three real-world networks.

dataset	nodes	edges	radius	diameter	clustering coefficient
FRIEND	1858	12534	7	14	0.14
EMAIL	1005	25571	4	7	0.40
BIBLE	1773	16401	5	8	0.72

6 Conclusion and Outlook

This paper proposes and formalizes the social capital-based network building (SCNB) problem and develops a Q-learning algorithm FQL for it that strikes a

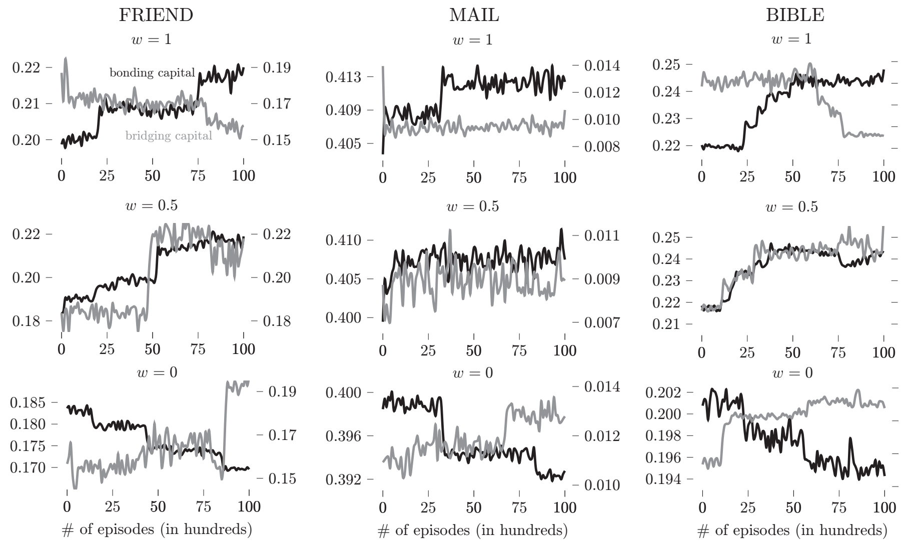

Fig. 4. Results of bonding (left axis and dark lines) and bridging capital (right axis and gray lines) in three real-world networks during 10000 episodes with varying preference weight w. Each curve is averaged over 10 independent runs.

balance between efficiency and performance. The experimental results show that the performance of learning efficiency is relevant to the structural properties of networks. A highly centralized structure can significantly accelerate learning speed and effect.

Ideas and methods proposed in this paper represent a novel research initiative. There are several potential directions for future work. A fairly straightforward expansion is to explore learning algorithms to gain social capital in evolving networks. Another future challenge is to employ representative learning and deep reinforcement learning to our model for predigesting the observation and transferring learned knowledge to other networks. A third future work is to investigate the emergence of social patterns when multiple agents simultaneously learn to gain social capital.

References

1. Alaa, A.M., Ahuja, K., van der Schaar, M.: A micro-foundation of social capital in evolving social networks. IEEE Transactions on Network Science and Engineering **5**(1), 14–31 (2018)
2. Aldrich, D.P., Meyer, M.A.: Social capital and community resilience. American behavioral scientist **59**(2), 254–269 (2015)
3. Barabási, A.L., Albert, R.: Emergence of scaling in random networks. science **286**(5439), 509–512 (1999)

4. Bouma, J., Bulte, E., van Soest, D.: Trust and cooperation: Social capital and community resource management. Journal of environmental economics and management **56**(2), 155–166 (2008)
5. Bourdieu, P.: The forms of capital. Handbook of Theory and Research for the Sociology of Education (1986)
6. Brânzei, S., Larson, K.: Social distance games. In: The 10th International Conference on Autonomous Agents and Multiagent Systems-Volume 3. pp. 1281–1282. International Foundation for Autonomous Agents and Multiagent Systems (2011)
7. Burt, R.S.: Structural holes and good ideas. American journal of sociology **110**(2), 349–399 (2004)
8. Coleman: Foundations of social theory. Harvard university press (1994)
9. Coleman, J.S.: Social capital in the creation of human capital. American journal of sociology **94**, S95–S120 (1988)
10. Erdos, P., Rényi, A.: On the evolution of random graphs. Publ. Math. Inst. Hung. Acad. Sci **5**(1), 17–60 (1960)
11. Girvan, M., Newman, M.E.: Community structure in social and biological networks. Proceedings of the national academy of sciences **99**(12), 7821–7826 (2002)
12. Granovetter, M.S.: The strength of weak ties. In: Social networks, pp. 347–367. Elsevier (1977)
13. Hällsten, M., Edling, C., Rydgren, J.: Social capital, friendship networks, and youth unemployment. Social Science Research **61**, 234–250 (2017)
14. Jackson, M.O.: A survey of network formation models: stability and efficiency. Group formation in economics: Networks, clubs, and coalitions **664**, 11–49 (2005)
15. Lin, N.: Social capital: A theory of social structure and action, vol. 19. Cambridge university press (2002)
16. Moskvina, A., Liu, J.: How to build your network? a structural analysis. International Joint Conference on Artificial Intelligence (IJCAI-2016) pp. 2597–2603 (2016)
17. Myers, D.: Relationship rewards. Social psychology pp. 392–439 (2010)
18. Page, L., Brin, S., Motwani, R., Winograd, T.: The pagerank citation ranking: Bringing order to the web. Tech. rep., Stanford InfoLab (1999)
19. Putnam, R.D.: Bowling alone: The collapse and revival of American community. Simon and Schuster (2001)
20. Rossi, R.A., Ahmed, N.K.: The network data repository with interactive graph analytics and visualization. In: Proceedings of the Twenty-Ninth AAAI Conference on Artificial Intelligence (2015)
21. Skyrms, B., Pemantle, R.: A dynamic model of social network formation. In: Adaptive networks, pp. 231–251. Springer (2009)
22. Sutton, R.S., Barto, A.G.: Reinforcement learning: An introduction. MIT press (2018)
23. Tong, H., Faloutsos, C., Pan, J.Y.: Fast random walk with restart and its applications. In: Sixth International Conference on Data Mining (ICDM'06). pp. 613–622. IEEE (2006)
24. Watkins, C.J., Dayan, P.: Q-learning. Machine learning **8**(3-4), 279–292 (1992)
25. Watts, D.J., Strogatz, S.H.: Collective dynamics of 'small-world'networks. nature **393**(6684), 440 (1998)
26. Yan, B., Chen, Y., Liu, J.: Dynamic relationship building: Exploitation versus exploration on a social network (2017)
27. Yan, B., Liu, Y., Liu, J., Cai, Y., Su, H., Zheng, H.: From the periphery to the core: Information brokerage in an evolving network. In: International Joint Conference on Artificial Intelligence (IJCAI-2018). pp. 3912–3918 (2018)