Social Structure Emergence: A Multi-agent Reinforcement Learning Framework for Relationship Building

Extended Abstract

Yang Chen, Jiamou Liu
The University of Auckland
Auckland, New Zealand
{yang.chen,jiamou.liu}@auckland.ac.nz

He Zhao, Hongyi Su
Beijing Institute of Technology
Beijing, China
{2120171104,henrysu}@bit.edu.cn

ABSTRACT

Social structures naturally arise from social networks, yet no model well interprets the emergence of structural properties in a unified dimension. Here, we unify explanations for the emergence of network structures by revealing the pivotal role of *social capital*, i.e., benefits that a society grants to individuals, in network formation. We propose a game-based framework *social capital games* that mathematically conceptualizes social capital. Through this framework, individuals are regarded as independent learning agents that aim to gain social capital via building interpersonal ties. We adopt multi-agent reinforcement learning (MARL) to train agents. By varying configurations of the game, we observe the emergence of classical structures of community, small-world, and core-periphery.

KEYWORDS

Network formation; multi-agent reinforcement learning; network structure; relationship building

ACM Reference Format:
Yang Chen, Jiamou Liu, He Zhao and Hongyi Su. 2020. Social Structure Emergence: A Multi-agent Reinforcement Learning Framework for Relationship Building. In *Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), Auckland, New Zealand, May 9–13, 2020,* IFAAMAS, 3 pages.

1 INTRODUCTION

Numerous real-life social networks exhibit prominent structural properties. Take, as an example, co-authorship networks that show *community* structure, where scholars in the same research field form a collaboration group [10]. Another example is *small-world* that is often observed in online social networks, where any two users are connected through a few intermediate acquaintances [26]. A third example is *core-periphery*, where a core sits in the center, while others stay at the outskirts [7]. Uncovering emergence of social structures can bring insights into how social networks form, function and evolve. However, no theory yet achieves a unified interpretation of the natural emergence of social structures.

Existing works on *network formation* aim to explain the emergence of social structures. Traditional approaches to network formation fall into two main paradigms: *random events*-based and *strategic decisions*-based. Random events-based models generate networks with ad-hoc designs that mimic real-world networks [1, 14, 17], but neglect agents' behavioral acquisitions. Strategic decisions-based models provide explanations for how social structures emerge as equilibria of *network formation games* [2, 13, 15, 21]. However, they are one-shot models and thus neglect dynamics of networks.

Recent advances in multi-agent reinforcement learning (MARL) and deep learning and sparks a new research perspective for problems with social concerns [12, 18, 20, 23, 24]. In this paper, we propose a game-based and MARL-centered framework, *social capital games* (SCGs), that aims to unify the explanation for social structure emergence. Social capital has been shown to be tightly correlated with social structures [4]. Thus, we define utilities in SCGs as social capital. We adopt multi-agent reinforcement learning (MARL) to train agents. By varying configurations, we reproduce the emergence of three aforementioned classical social structures.

2 MODEL

Dynamic Networks. Let $N = \{1, 2, \dots, n\}$ be a finite set of agents. We define the complete graph g^N as the set of all subsets of N of size 2. Hence $\{g \mid g \subseteq g^N\}$ denotes the set of all possible graphs on N. For any $i \neq j$, we write $ij \in g$ to denote an undirected edge between i and j. To capture the creation of links, for any $g' \subseteq g^N$, let $g \cup g'$ denote the integrated graph obtained via adding each link $ij \in g'$ into g. Denoted by $\mathcal{N}_d(i)$, the *d-hop neighbor set* of i are the set of i's neighbors with distance d. We denote $\mathcal{N}_d[i] := \{j \in \mathcal{N}_x(i) \mid x \leq d\}$ all i's neighbors within distance d. We assume an agent i only has local information, i.e., 2-hop neighbors $o_i := \{jk \mid j, k \in \mathcal{N}_2[i]\}$. A *dynamic network* is a sequence of graphs $G = g^0, g^1, \dots, g^\ell$ that evolves in finite discrete time $0, 1, \dots, \ell$, where ℓ is the *termination step*. Throughout, we use superscript t and subscript i to denote the corresponding notation derived from time step t and agent i, respectively. For $t < \ell$, each agent $i \in N$ builds a link to another agent a_i^t from o_i^t. All agents make decisions simultaneously. Formally, $\forall 0 \leq t < \ell : g^{t+1} = g^t \cup \{ia_i^t\}_{i \in N}$, where $a_i^t \in \mathcal{N}_2^t(i)$.

Social Capital. A well-known dichotomy defines two types of social capital: *bonding capital*, which refers to welfare such as trust and norms [3], and *bridging capital*, which amounts to benefits in terms of influence and power [5]. • We adopt the formalization of bonding capital as in [6], which uses *personalized PageRank index* to capture benefits rising from neighbors. The metric is adapted from PageRank to capture the likelihood of a random walk from i (with restart) that reaches j [22]. Let PageRank index pr_j denote the probability that node j is accessed after convergence of the walk. The bonding capital of i is defined by summing personalized PageRank indices between i and i's neighbors, i.e., $\mathrm{bo}_i := \sum_{j \in \mathcal{N}_1(i)} \mathrm{pr}_j$. • The formalization of bridging capital is straightforward by using *betweenness centrality* [1]: $\mathrm{br}_i := \sum_{j \neq i \neq k \in N} \sigma_{jk}(i)/\sigma_{jk}$, where

Proc. of the 19th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2020), B. An, N. Yorke-Smith, A. El Fallah Seghrouchni, G. Sukthankar (eds.), May 9–13, 2020, Auckland, New Zealand. © 2020 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

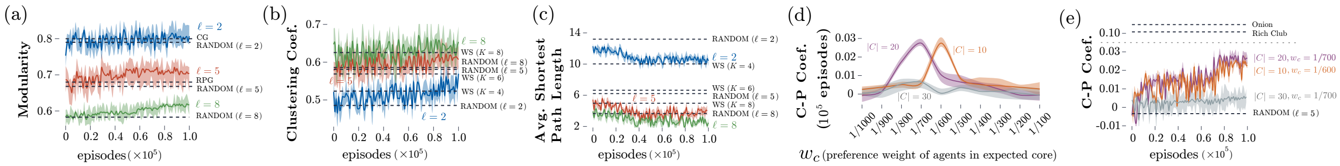

Figure 1: Results for (a) modularity; (b) clustering coefficients; (c) average shortest path length; (d) C-P coefficients after 10^5 episodes; (e) C-P coefficients under the configurations of three peaks as shown in (d). Results are averaged over 10 independent runs. Parameter settings: graph embedding iterations $T = 4$, vector dimension $p = 32$, minibatch size $b = 32$.

σ_{jk} is the number of shortest paths between nodes j and k, and $\sigma_{jk}(i)$ is the number of shortest paths passing i. • As an agent may have different preferences to two types of capital, we employ a *preference weight* $w \in [0, 1]$ to define the *mixed capital*: $\mathrm{mix}_{i,w} := w\mathrm{bo}_i + (1 - w)\mathrm{br}_i$.

Social Capital Games (SCGs). A *social capital game* is a tuple (N, W, g^0, ℓ), where $N = \{1, 2, \ldots, n\}$ is a finite set of agents; $W = (w_1, w_2, \ldots, w_{|N|})$ is a *preference vector*, in which each entry w_i is i's preference weight; $g^0 \subseteq g^N$ is the *initial network*; $\ell \in \mathbb{N}^+$ is the *termination step*. Conceptually, one can view an SCG as a multi-stage game played with imperfect information. We measure the *immediate utility* of an agent as the increment of the mix capital between two consecutive time steps: $u_i^{t+1} := \mathrm{mix}_{i,w_i}^{t+1} - \mathrm{mix}_{i,w_i}^{t}$. A *policy* of an agent $i \in N$ is a function π_i defined on all possible 2-hop neighbors of i such that $\pi_i(o_i) = a \in \mathcal{N}_2(i)$ for any $o_i \subseteq g^N$. The goal of agent i is to find a policy that maximizes the *cumulative utility* $U_i^\ell := \sum_{i=1}^{t} u_i^t$. Thus, an underlying dynamic network $G = g^0, g^1, \ldots, g^\ell$ is generated. By repeating the game, the trajectory of g^ℓ represents the evolution of social structures.

3 MARL FOR SOCIAL CAPITAL GAMES

Our learning method is adapted from S2V-DQN as in [16], which incorporates *graph embedding* and RL to solve combinatorial problems on graphs. In our proposed MARL method for SCGs, all agents independently and synchronously use S2V-DQN to learn a policy. Each agent $i \in N$ estimates the quality of linking to another agent $a \in \mathcal{N}_2(i)$ under o_i using an *evaluation function* $Q_i(o_i, a)$. The policy π_i functions greedily w.r.t. Q_i, i.e.,

$$\pi_i(o_i) := \arg\max_{a \in \mathcal{N}_2(i)} Q_i(o_i, a). \tag{1}$$

S2V-DQN uses `structure2vec` [8] to parameterize $Q_i(o_i, a; \Theta_i)$ that computes a p-dimensional feature embedding $\boldsymbol{\mu}_j$ for each node j involved in o_i. $\boldsymbol{\mu}_j$ is iteratively updated. Initialized as $\mathbf{0}$, after T iterations, $\boldsymbol{\mu}_j$ will contain information about its T-hop neighbors as determined by the structure of o_i. The update rule is:

$$\boldsymbol{\mu}_j^{(t+1)} = \mathrm{ReLU}\left(\boldsymbol{\theta}_1 \boldsymbol{x}_j + \boldsymbol{\theta}_2 \sum\nolimits_{k \in \mathcal{N}_1(j)} \boldsymbol{\mu}_k^{(t)}\right), \tag{2}$$

where $\boldsymbol{\theta}_1 \in \mathbb{R}^{p \times 2}, \boldsymbol{\theta}_2 \in \mathbb{R}^{p \times p}$ are model parameters and ReLU is the rectified linear unit ($\mathrm{ReLU}(z) = \max(0, z)$). The vector $\boldsymbol{x}_j$ incorporates explicit features of j. Here, we set $\boldsymbol{x}_j = (w_j, \mathrm{dist}(i, j))^\intercal$.

The embedding $\boldsymbol{\mu}_a$ and the pooled embedding over o_i, $\boldsymbol{\phi}(o_i) := \sum_{j \in \mathcal{N}_2[i]} \boldsymbol{\mu}_j$, are used as the surrogates for a and o_i, resp., i.e.,

$$Q_i(o_i, a; \Theta_i) = \boldsymbol{\theta}_3^\intercal \mathrm{relu}(\boldsymbol{\theta}_4 \boldsymbol{\phi}(o_i) \ \oplus \ \boldsymbol{\theta}_5 \boldsymbol{\mu}_a), \tag{3}$$

where $\boldsymbol{\theta}_3 \in \mathbb{R}^{2p}, \boldsymbol{\theta}_4, \boldsymbol{\theta}_5 \in \mathbb{R}^{p \times p}$, and $\oplus$ is the concatenation operator. The parameterized evaluation function $Q_i(o_i, a; \Theta_i)$ is based on a collection of 5 parameters $\Theta_i = \{\boldsymbol{\theta}_m\}_{1 \le m \le 5}$, which will be learned. The Q-learning is used to learn Θ_i and the *experience replay* is used to update Θ_i. For each step, a minibatch of tuples (size of b) is randomly sampled from *experience dataset* $\mathcal{D}_i$. Then stochastic gradient descent is executed on the following squared loss:

$$\mathcal{L}(\Theta_i) = \mathbb{E}_{(o,a,r,o') \sim \mathcal{D}_i}\left[(y - Q_i(o, a; \Theta_i))^2\right], \tag{4}$$

where $y = r + \max_{a'} Q_i(o', a'; \Theta_i)$ is the update target.

4 EMERGENCE OF SOCIAL STRUCTURES

General Settings. We train $|N| = 100$ agents and set the initial network g^0 as a regular ring lattice. We vary termination step $\ell \in \{2, 5, 8\}$. We use *modularity* [19], *clustering coefficient* and *average shortest path length* [26], and *C-P coefficient* [11] to measure the significance of community, small-world and core-periphery, resp.
Baselines. For each ℓ, we generate 100 random networks for reference, where in each step each agent randomly links to an agent. We also adopt network generation models as baselines: • *Caveman graphs* (CG) [25] and *Random partition graphs* (RPG) [9] for community. • *Watts-Strogatz* (WS) model [26] (start from a regular lattice, each node connected to K neighbors and $K/2$ on each side. Then edges are randomly rewired with probability $p = 0.01$.) for small-world. • *rich club model* and *onion model* [7] for core-periphery.
Configurations. The intuitions and configurations are listed as follows: • Community emerges when all agents are in pure pursuit of bonding capital. We set preference weight $w_i = 1$ for all $1 \le i \le 100$. • Small-world emerges when all agents are in pure pursuit of bridging capital. We set $w_i = 0$ for all $1 \le i \le 100$. • Core-periphery emerges when a group of agents are in pure pursuit of bonding capital, while the remaining agents show mixed preferences to bonding and bridging capital. We randomly select a subset $C \subset N$ (expected core) with varying size in $\{10, 20, 30\}$. For all $c \in C$, we vary w_c from 1/1000 to 1/100. For all $p \in N \backslash C$ (expected periphery), we set $w_p = 1$. Throughout, we fix ℓ to an intermediate value, 5.
Results. The statistical information of g^ℓ by our proposed framework and baselines is plotted in Fig. 1. • Modularity grows and fluctuates at a high level as the number of episodes grows, compared to baselines. • Our framework achieves comparable high clustering coefficients and lower average shortest path lengths for each value of ℓ. • Three peaks of C-P coefficients occur under $(|C|, w_c) = (10, 1/600)$, $(20, 1/700)$ and $(30, 1/700)$. Our framework outputs accepted high C-P coefficients compared to baselines. Overall, our proposed framework successfully reproduce and explain the emergence of various classical social network structures.

REFERENCES

[1] Ahmed M Alaa, Kartik Ahuja, and Mihaela van der Schaar. 2017. A microfoundation of social capital in evolving social networks. *IEEE Transactions on Network Science and Engineering* 5, 1 (2017), 14–31.

[2] Francis Bloch and Matthew O Jackson. 2006. Definitions of equilibrium in network formation games. *International Journal of Game Theory* 34, 3 (2006), 305–318.

[3] Pierre Bourdieu. 1986. The forms of capital. *Handbook of Theory and Research for the Sociology of Education* (1986).

[4] Ronald S Burt. 2000. The network structure of social capital. *Research in organizational behavior* 22 (2000), 345–423.

[5] Ronald S Burt. 2004. Structural holes and good ideas. *American journal of sociology* 110, 2 (2004), 349–399.

[6] Yijin Cai, Hong Zheng, Jiamou Liu, Bo Yan, Hongyi Su, and Yiping Liu. 2018. Balancing the Pain and Gain of Hobnobbing: Utility-Based Network Building over Atributed Social Networks. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. 193–201.

[7] Peter Csermely, András London, Ling-Yun Wu, and Brian Uzzi. 2013. Structure and dynamics of core/periphery networks. *Journal of Complex Networks* 1, 2 (2013), 93–123.

[8] Hanjun Dai, Bo Dai, and Le Song. 2016. Discriminative embeddings of latent variable models for structured data. In *International conference on machine learning (ICML)*. 2702–2711.

[9] Santo Fortunato. 2010. Community detection in graphs. *Physics reports* 486, 3-5 (2010), 75–174.

[10] Michelle Girvan and Mark EJ Newman. 2002. Community structure in social and biological networks. *Proceedings of the national academy of sciences* 99, 12 (2002), 7821–7826.

[11] Petter Holme. 2005. Core-periphery organization of complex networks. *Physical Review E* 72, 4 (2005), 046111.

[12] Edward Hughes, Joel Z Leibo, Matthew Phillips, Karl Tuyls, Edgar Dueñez-Guzman, Antonio García Castañeda, Iain Dunning, Tina Zhu, Kevin McKee, Raphael Koster, et al. 2018. Inequity aversion improves cooperation in intertemporal social dilemmas. In *Advances in neural information processing systems (NIPS)*. 3326–3336.

[13] Matthew O Jackson. 2005. A survey of network formation models: stability and efficiency. *Group formation in economics: Networks, clubs, and coalitions* 664 (2005), 11–49.

[14] Matthew O Jackson and Brian W Rogers. 2007. Meeting strangers and friends of friends: How random are social networks? *American Economic Review* 97, 3 (2007), 890–915.

[15] Matthew O Jackson and Asher Wolinsky. 1996. A strategic model of social and economic networks. *Journal of economic theory* 71, 1 (1996), 44–74.

[16] Elias Khalil, Hanjun Dai, Yuyu Zhang, Bistra Dilkina, and Le Song. 2017. Learning combinatorial optimization algorithms over graphs. In *Advances in Neural Information Processing Systems (NIPS)*. 6348–6358.

[17] Jure Leskovec, Lars Backstrom, Ravi Kumar, and Andrew Tomkins. 2008. Microscopic evolution of social networks. In *Proceedings of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 462–470.

[18] Mihail Mihaylov, Karl Tuyls, and Ann Nowé. 2014. A decentralized approach for convention emergence in multi-agent systems. *Journal of Autonomous Agents and Multi-Agent Systems (JAAMAS)* 28, 5 (2014), 749–778.

[19] Mark EJ Newman. 2006. Modularity and community structure in networks. *Proceedings of the national academy of sciences* 103, 23 (2006), 8577–8582.

[20] Sandip Sen and Stéphane Airiau. 2007. Emergence of norms through social learning.. In *20th International Joint Conference on Artificial Intelligence (IJCAI)*, Vol. 1507. 1512.

[21] Yangbo Song and Mihaela van der Schaar. 2015. Dynamic network formation with incomplete information. *Economic Theory* 59, 2 (2015), 301–331.

[22] Hanghang Tong, Christos Faloutsos, and Jia-Yu Pan. 2006. Fast random walk with restart and its applications. In *6th International Conference on Data Mining (ICDM)*. IEEE, 613–622.

[23] Daniel Villatoro, Jordi Sabater-Mir, and Sandip Sen. 2011. Social instruments for robust convention emergence. In *Proceedings of the 22nd International Joint Conference on Artificial Intelligence (IJCAI)*.

[24] Jane X Wang, Edward Hughes, Chrisantha Fernando, Wojciech M Czarnecki, Edgar A Duéñez-Guzmán, and Joel Z Leibo. 2019. Evolving intrinsic motivations for altruistic behavior. In *Proceedings of the 18th International Conference on Autonomous Agents and MultiAgent Systems (AAMAS)*. 683–692.

[25] Duncan J Watts. 1999. Networks, dynamics, and the small-world phenomenon. *American Journal of sociology* 105, 2 (1999), 493–527.

[26] Duncan J Watts and Steven H Strogatz. 1998. Collective dynamics of 'small-world' networks. *Nature* 393, 6684 (1998), 440.