Maximizing Influence-Based Group Shapley Centrality

Extended Abstract

Ruben Becker
Gran Sasso Science Institute
L'Aquila, 67100, Italy
ruben.becker@gssi.it

Gianlorenzo D'Angelo
Gran Sasso Science Institute
L'Aquila, 67100, Italy
gianlorenzo.dangelo@gssi.it

Hugo Gilbert
Université Paris-Dauphine, Université PSL, CNRS, LAMSADE
75016 Paris, France
hugo.gilbert@dauphine.psl.eu

ABSTRACT

A key problem in network analysis is the influence maximization problem, which consists of finding a set S of at most k seed users in a social network, such that the spread of information from S is maximized. We investigate the problem of choosing the best set of seeds when there exists an unknown pre-existing set of seed nodes. Our work extends the one of Chen and Teng (WWW'17) who introduced the so-called Shapley centrality of a node to measure the efficiency of nodes acting as seeds within a pre-existing but unknown set of seeds. We instead consider the question: Which set of cardinality k to target in this kind of scenario? The resulting optimization problem reveals very challenging, that is, assuming common computational complexity conjectures, we obtain strong hardness of approximation results. Nevertheless, we design a greedy algorithm which achieves an approximation factor of $\frac{1-1/e}{k} - \epsilon$ for any $\epsilon > 0$, showing that not all is lost in settings where k is bounded.

KEYWORDS

Influence Maximization; Group-Shapley Value

ACM Reference Format:
Ruben Becker, Gianlorenzo D'Angelo, and Hugo Gilbert. 2021. Maximizing Influence-Based Group Shapley Centrality: Extended Abstract. In *Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), Online, May 3–7, 2021*, IFAAMAS, 3 pages.

1 MOTIVATION

Node centrality and spread of information or influence are two main topics in network analysis. The former regards the problem of determining the most important nodes in a network according to some measure of importance, while the latter studies mathematical models to represent how information spreads in a Social Network (SN) in which nodes are able to communicate with each other.

In order to measure the centrality of nodes in a network a real-valued function, called centrality index, associates a real number with each node that reflects its importance or criticality within the network. Chen and Teng [4] initiated the study of the interplay between spreading dynamics and network centrality by defining two centrality indices based on dynamic models for influence diffusion: the *single node influence centrality*, which measures the centrality of a node by its capability of spreading information when acting alone, and the *Shapley centrality*, which uses the Shapley value to measure the capability of a node to increase the spreading capacity of a pre-existing but unknown group of nodes. The Shapley value – a well-known concept from cooperative game theory – models this scenario by assessing the expected relevance of each player within a *random* subset of players (also called coalition), where the expectation is taken over the possible coalitions. More formally, given a characteristic function τ that maps each coalition to the total payoff that this coalition receives, the Shapley value of a player i is defined as the expected payoff that i adds to any coalition, w.r.t. the function τ. The Shapley centrality index studied by Chen and Teng [4] measures the centrality of a node by using the Shapley value and the spreading function σ as characteristic function.

Most centrality indices neglect the relevance that coalitions of individuals and their coordination play in SNs. For this reason, many centrality indices have been generalized to *group centrality indices* which are real-valued functions over subsets of nodes instead of single nodes. A group centrality index is fundamentally different from a combination of the individual centrality indices of the nodes in the group, as it captures the relevance of the set as a whole. This work extends Chen and Teng's notion of influence-based Shapley centrality from single nodes to *groups of nodes* by using the concept of the Group Shapley value. Our Influence-based Group Shapley (IGS) centrality associates to a set S of nodes, the expected gain in influence that S adds to a random pre-existing seed set T. We investigate the problem of finding a set S of cardinality k with highest IGS value. We believe that this way of evaluating the importance of a seed set is of high practical interest. Assume an entity wants to spread a piece of news, while having a budget to influence k users, at the same time knowing that already some users are aware of the information and will spread it anyhow. The central entity, however, may have no knowledge about who these users are. In this case, it should target a set of seed users with large IGS value.[1]

2 PRELIMINARIES

The Group Shapley Value. In cooperative game theory [3, 10], a game on $n \geq 2$ players is commonly formalized by a characteristic function $\tau : 2^{[n]} \to \mathbb{R}$ that assigns to every subset $S \subseteq [n]$ of players, also called a coalition, a value $\tau(S)$. The *Group Shapley value* [5, 9], for a subset $S \subseteq [n]$ of players in a game τ is defined as

$$\phi_\tau^{\mathrm{Sh}}(S) := \sum_{T \subseteq [n] \setminus S} \frac{|T|!(n - |S| - |T|)!}{(n - |S| + 1)!} \cdot (\tau(T \cup S) - \tau(T)),$$

i.e., the Group Shapley value quantifies how much S is able to add in the game τ to a random subset T that is generated as follows: We first sample an integer $t \in \{0, \dots, n - |S|\}$ uniformly at random and then pick a set T of size t in $[n] \setminus S$ uniformly at random.

[1] See [1] for the complete proofs of all results reported on in this extended abstract.

Proc. of the 20th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2021), U. Endriss, A. Nowé, F. Dignum, A. Lomuscio (eds.), May 3–7, 2021, Online.
© 2021 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). All rights reserved.

Influence Maximization (IM). We are interested in the Group Shapley value for functions that describe information propagation in SNs. Two of the most popular models for describing such information propagation are the *Independent Cascade* (IC) and *Linear Threshold* (LT) models [7]. In both models, we are given a directed graph $G = (V, E)$ where V is a set of n nodes, values $\{p_{uv} \in [0, 1] : (u, v) \in E\}$ and an initial node set $A \subseteq V$ called *seed set.* A spread of influence from A is defined as a randomly generated sequence of node sets $(A_t)_{t \in \mathbb{N}}$, where $A_0 = A$ and $A_{t-1} \subseteq A_t$. These sets represent active users, i.e., a node v is *active* at time step t if $v \in A_t$. We say that the sequence converges as soon as $A_{t^*} = A_{t^*+1}$, for some time step $t^* \geq 0$ called the time of quiescence. For a set A, we use $\sigma(A) = \mathbb{E}[|A_{t^*}|]$ to denote the expected number of nodes activated at the time of quiescence when running the process with seed set A. In IM, the standard objective is to find a set A maximizing $\sigma(A)$ under a cardinality constraint. Both the IC and LT model are special cases of the more general *Triggering Model*, see [7, Proofs of Theorem 4.5 and 4.6] that we consider in this work.

Influence-based Group Shapley Centrality. Chen and Teng [4] consider the Shapley value of nodes w.r.t. the influence spread function σ in a SN modeled by the Triggering Model. They use the resulting *Shapley centrality* $\phi_\sigma^{\mathrm{Sh}}(i)$ as a measure of centrality of node i. In this work, we consider $\phi_\sigma^{\mathrm{Sh}}$, i.e., the Group Shapley value w.r.t. σ. We call this value the *Influence-based Group Shapley* (IGS) centrality of S. We refer to it as $\phi^{\mathrm{Sh}}(S)$ omitting σ as an index.

The central problem of this paper consists in finding sets S of size at most k that have large IGS centrality among all such sets. We call this the Max-Shapley-Group problem. The most important tool for studying this problem are so-called RR sets [2, 12]. In fact, there exists a concise formulation of ϕ^{Sh} in terms of RR sets:

Lemma 2.1 (IGS centrality via RR sets). *For any $S \subseteq V$,*

$$\phi^{\mathrm{Sh}}(S) = n \cdot \mathbb{E}_R\Big[\frac{\mathbb{1}_{R \cap S \neq \emptyset}}{|R \setminus S| + 1}\Big].$$

3 HARDNESS OF APPROXIMATION

Unfortunately, Max-Shapley-Group in the IC model is, up to a constant factor, as hard to approximate as Densest-k-Subgraph.

Theorem 3.1. *Let $\alpha \in (0, 1]$. If there is an α-approximation algorithm for Max-Shapley-Group, then there is an $\alpha/8$-approximation algorithm for Densest-k-Subgraph.*

A number of strong hardness of approximation results are known for Densest-k-Subgraph: (1) Densest-k-Subgraph cannot be approximated within $1/n^{o(1)}$ if the Gap Exponential Time Hypothesis holds [8]. (2) Densest-k-Subgraph cannot be approximated within any constant if the Unique Games with Small Set Expansion conjecture holds [11]. (3) Densest-k-Subgraph cannot be approximated within $n^{-(\log \log n)^{-c}}$ for some constant c if the Exponential Time Hypothesis holds [8]. We provide a reduction which yields the same hardness results for Max-Shapley-Group. In particular, according to (1) and our reduction, it is unlikely to find anything better than an (n^{-c})-approximation for Max-Shapley-Group, where c is a constant. Furthermore, for all settings where $k = O(n^c)$, such an algorithm is implied by our result in Section 4.

4 A SIMPLE APPROXIMATION ALGORITHM

A good approximation result, as for example a constant-factor approximation, is unlikely for the Max-Shapley-Group problem. We however obtain a positive result for small values of k.

Approximating ϕ^{Sh} through RR sets. By sampling a sufficient number of RR sets, we can give a set function $\hat{\phi}^{\mathrm{Sh}}$ that with high probability approximates ϕ^{Sh} to within a factor of $1 \pm \epsilon$. The idea is to approximate the expected value in Lemma 2.1 via a Chernoff bound. This is captured by the following lemma.

Lemma 4.1. *Let $\epsilon \in (0, 1)$ and let $R_1, \ldots, R_t$ be a sequence of t RR sets. For a value of t that is polynomial in n and ϵ^{-1}, w.h.pr.,*

$$\hat{\phi}^{\mathrm{Sh}}(S) := \frac{n}{t} \sum_{i=1}^{t} \frac{\mathbb{1}_{R_i \cap S \neq \emptyset}}{|R_i \setminus S| + 1}$$

is a $(1 \pm \epsilon)$-approximation of ϕ^{Sh}.

The Harmonic-Max-Hitting-Set problem. Lemma 4.1 suggests to sample a near-linear number t of RR sets and compute a set of nodes S that maximizes $\hat{\phi}^{\mathrm{Sh}}(S)$. We call the resulting problem the Harmonic-Max-Hitting-Set problem: the input consists of a set $X = \{x_1, \ldots, x_n\}$, a set $Z = \{Z_1, \ldots, Z_m\}$ of subsets of X, and an integer k. The task is to find a subset $S \subseteq X$ s.t. $|S| \leq k$ maximizing

$$f_Z(S) := \sum_{i=1}^{m} \frac{\mathbb{1}_{Z_i \cap S \neq \emptyset}}{|Z_i \setminus S| + 1}.$$

This is a non-linear variant of the well-known Max-Hitting-Set problem (itself equivalent to the Max-Set-Cover problem [6]) in which the objective function is $\sum_{i=1}^{m} \mathbb{1}_{Z_i \cap S \neq \emptyset}$. The problem of maximizing $\hat{\phi}^{\mathrm{Sh}}$ can be stated as a Harmonic-Max-Hitting-Set problem by letting $X = V$ and Z be the set of generated RR sets.

Approximation Algorithm. Consider an instance (X, Z, k) and define the following set function $h_Z(S) := \sum_{i=1}^{m} \mathbb{1}_{Z_i \cap S \neq \emptyset}/|Z_i|$. Note the similarity between h_Z and f_Z. The approximation algorithm that we propose is to greedily maximize h_Z instead of f_Z. Why would this be a good idea? (1) The set function h_Z is monotone and submodular; thus the greedy algorithm will yield a $1 - 1/e$ approximation to maximizing h_Z. (2) Given a set $S \subseteq X$ with $|S| \leq k$, it holds that $f_Z(S) \geq h_Z(S) \geq f_Z(S)/k$, that is, the error when passing from h_Z to f_Z is at most k. Hence, if we denote by S_f^* (resp. S_h^*) an optimal solution of size k for maximizing f_Z (resp. h_Z), we have that $h_Z(S_h^*) \geq h_Z(S_f^*) \geq f_Z(S_f^*)/k$. Now let S be the solution of size k returned by the greedy algorithm. Then, S is a $(1 - 1/e)/k$ approximation to maximizing f_Z as

$$f_Z(S) \geq h_Z(S) \geq \Big(1 - \frac{1}{e}\Big) \cdot h_Z(S_h^*) \geq \frac{1 - 1/e}{k} \cdot f_Z(S_f^*).$$

Theorem 4.2. *Let $\epsilon \in (0, 1)$. There is an algorithm with running time polynomial in n and ϵ^{-1} that, with high probability, returns a $\frac{1-1/e}{k} - \epsilon$-approximation for the Max-Shapley-Group problem.*

ACKNOWLEDGMENTS

This work was partially supported by the Italian MIUR PRIN 2017 Project "ALGADIMAR" Algorithms, Games, and Digital Markets.

REFERENCES

[1] Ruben Becker, Gianlorenzo D'Angelo, and Hugo Gilbert. 2020. Maximizing Influence-based Group Shapley Centrality. *CoRR* abs/2003.07966 (2020).

[2] Christian Borgs, Michael Brautbar, Jennifer T. Chayes, and Brendan Lucier. 2014. Maximizing Social Influence in Nearly Optimal Time. In *Proceedings of the 25th Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2014, Portland, Oregon, USA, January 5-7, 2014*. 946–957.

[3] Georgios Chalkiadakis, Edith Elkind, and Michael Wooldridge. 2011. Computational aspects of cooperative game theory. *Synthesis Lectures on Artificial Intelligence and Machine Learning* 5, 6 (2011), 1–168.

[4] Wei Chen and Shang-Hua Teng. 2017. Interplay between social influence and network centrality: a comparative study on shapley centrality and single-node-influence centrality. In *Proceedings of the 26th international conference on World Wide Web (WWW)*. 967–976.

[5] Ramón Flores, Elisenda Molina, and Juan Tejada. 2019. Evaluating groups with the generalized Shapley value. *4OR* 17, 2 (2019), 141–172.

[6] Michael R. Garey and David S. Johnson. 2009. *Computers and intractability: a guide to the theory of NP-completeness* (27. print ed.). Freeman, New York.

[7] David Kempe, Jon M. Kleinberg, and Éva Tardos. 2015. Maximizing the Spread of Influence through a Social Network. *Theory of Computing* 11 (2015), 105–147.

[8] Pasin Manurangsi. 2017. Almost-polynomial ratio ETH-hardness of approximating densest k-subgraph. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*. 954–961.

[9] Jean-Luc Marichal, Ivan Kojadinovic, and Katsushige Fujimoto. 2007. Axiomatic characterizations of generalized values. *Discrete Applied Mathematics* 155, 1 (2007), 26–43.

[10] Roger B. Myerson. 2013. *Game theory*. Harvard university press.

[11] Prasad Raghavendra and David Steurer. 2010. Graph expansion and the unique games conjecture. In *Proceedings of the 42nd ACM Symposium on Theory of Computing, STOC 2010, Cambridge, Massachusetts, USA, 5-8 June 2010*. 755–764.

[12] Youze Tang, Xiaokui Xiao, and Yanchen Shi. 2014. Influence maximization: near-optimal time complexity meets practical efficiency. In *International Conference on Management of Data, SIGMOD 2014, Snowbird, UT, USA, June 22-27, 2014*. 75–86.