BECOMING GATEKEEPERS TOGETHER WITH ALLIES: COLLABORATIVE BROKERAGE OVER SOCIAL NETWORKS

YANG CHEN AND JIAMOU LIU
THE UNIVERSITY OF AUCKLAND, NEW ZEALAND
Email: jiamou.liu@auckland.ac.nz

INTRODUCTION

- **Social network integration** refers to a process where ties are created between disjoint groups of individuals.
- During the integration process, it is vital to establish key relationships which allow information to flow effectively.
- **Question:** Who should a team of individuals target to establish ties with in order to integrate into a network with the maximum effectiveness?
- The **network building problem** takes as input a graph G and aims to establish edges between a team of newcomers and a selection of nodes in G such that all nodes in the combined network are within certain distance-based bounds away from the newcomers.
- This problem has been investigated in many works e.g., [Moskvina, Liu 2016], [Yan, et al, 2018], when the "team of newcomers" contains only one element.
- We generalize these work to more than one newcomer and study this problem when the team of newcomers may have heterogeneous influencing power.

COLLAB-BROKERAGE PROBLEM

- A set of nodes $D \subseteq V$ is a **distance-ρ dominating (dom-ρ) set** if for all nodes $u \in V$, there is some $v \in D$ such that $\mathrm{dist}(v,u) \leq \rho$. Let $\delta_\rho(G)$ denote the size of a minimum dom-ρ set for G.
- For a directed graph $G=(V,E)$, a **distance-(ρ_1,ρ_2) dominating (dom-(ρ_1,ρ_2)) team** consists of a pair (D_1,D_2) of node sets where $D_1 \cap D_2 = \varnothing$, and for any node $u \in V$, either $\mathrm{dist}(v,u) \leq \rho_1$ for some $v \in D_1$ or $\mathrm{dist}(v,u) \leq \rho_2$ for some $v \in D_2$.
- An integer d is an **order** of a dom-(ρ_1,ρ_2) team (D_1,D_2) if $|D_2| \leq d$.
- Given a graph G and $d \in \mathbb{N}$, the **Collab-Brokerage** problem asks for a smallest order-d dom-(ρ_1,ρ_2) team.

THEORETICAL RESULTS

- It is NP-hard to approximate $\delta_1(G)$ on input graph of size N to within $(1-\alpha)\ln N$ [Moshkovitz 2012].
- The *Collab-Brokerage* problem can be solved in polynomial time over directed trees.

PROPOSED ALGORITHMS

Exact algorithm:

- DP: A dynamic programming algorithm that computes a smallest order-d dom-(ρ_1,ρ_2) team over directed trees in polynomial time.

Approximation algorithms:

- STDP-k: Approximate smallest (D_1,D_2) on general graphs by repeating DP on spanning trees for k trials. Take the best result as the final output.
- Greedy: Compute D_2 and D_1 sequentially w.r.t. a specific heuristic.
- REPL1: Firstly compute D_2 using Greedy by assuming $d=\infty$; Then place D_1 nodes to replace D_2 nodes in a purely greedy manner, until $|D_2|=d$.
- REPL2: Firstly compute D_1 using Greedy by assuming $d=0$; Then place D_2 nodes to replace and remove extra D_1 nodes in a purely greedy manner, until $|D_2|=d$.

Heuristics:

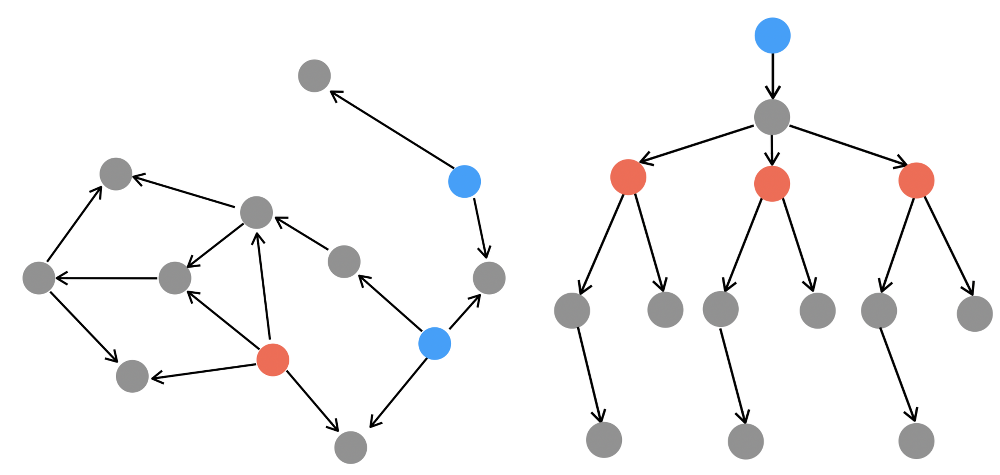
- Max: Select a node with maximum outdegree.
- Min: Assign priority to nodes with small indegrees. At each iteration, the heuristic adds to $D_2(D_1)$ a node that has distance at most $\rho_2(\rho_1)$.

EXAMPLES

•Left: A smallest order-1 dom-$(1,2)$ team on a directed graph, output by Greedy-Max.

•Right: A smallest order-3 dom-$(1,2)$ team on a directed tree, output by DP.

Reds nodes and blue nodes indicate D_2 and D_1, respectively.

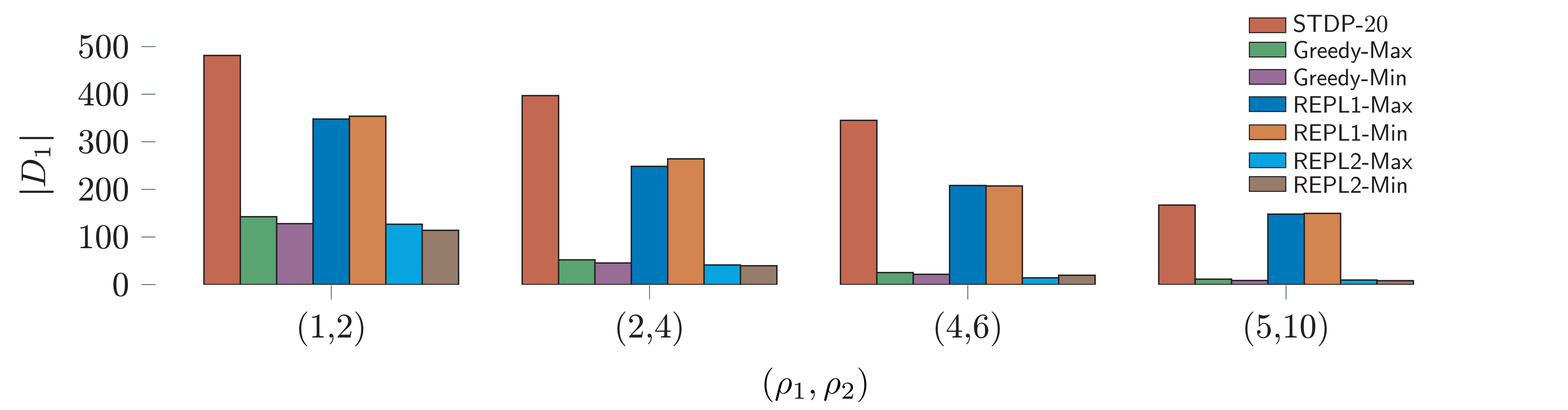


EXPERIMENTAL RESULTS

We conducted two types of experiments: **Test 1.** fix (ρ_1,ρ_2) to $(1,2)$ and vary d. **Test 2.** fix d to 50 and vary (ρ_1,ρ_2).

Experiment 1: Random Networks.

We tested approximation algorithms on three types of synthetic networks – directed Barabási Albert (BA) networks and Navigable Small World (NSW) networks. The figure shown below illustrates results for NSW networks.



Results.

We generated 10 NSW networks each initiated with a 32 by 32 grid (1024 nodes). Observe that

- STDP-k again performs worst among all algorithms.
- REPL2 surpasses REPL1 remarkably in both tests.
- For REPL2, the heuristic Min shows a slightly better performance than Max.
- Overall, we observe that among all categories of algorithms, replacement based algorithms (especially REPL2) show the averagely best performance on all synthetic networks. Moreover, the heuristic Max and Min should be chosen accordingly as the structural properties may vary over networks.

Experiment 2: Real-World Networks.

We conducted experiments on three datasets: (1) **Wiki-Vote (Wiki)** contains all the Wikipedia voting data from the inception of Wikipedia till January 2008. Nodes in the network represent Wikipedia users and a directed edge from i to j represents that i voted on j. (2) **Bitcoin OTC trust network (Bitcoin)** record anonymous Bitcoin trading on Bitcoin OTC with temporal information, where a directed edge ij denotes a trade between user i and user j. (3) **Cit-HepPh network (Cit)** is a high-energy physics citation network, which collects all papers from 1993 to 2003 on arXiv; A directed edge ij denotes that paper i cites j. The statistics of three real-world networks are summarized in the table below, where the minimum size of D_1 output by algorithms is highlighted in bold for each test.

		Wiki						**Bitcoin**						**Cit**					
		Greedy		REPL1		REPL2		Greedy		REPL1		REPL2		Greedy		REPL1		REPL2	
(ρ_1,ρ_2)	d	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
(1,2)	0	4821	**4812**	7035	7047	4821	**4812**	**2715**	3205	3650	5655	2725	3205	8892	**8327**	23005	22784	8892	**8327**
(1,2)	50	**4702**	**4702**	6975	6996	4747	4744	1672	2651	**1020**	3723	1202	1523	8055	7959	22324	22447	8542	**7898**
(1,2)	100	**4645**	4651	6924	6944	4694	4694	910	2571	**634**	3110	857	1136	7812	**7631**	21938	22289	8362	7692
(1,2)	150	**4592**	4601	6872	6894	4643	4644	715	2199	**472**	2580	695	952	7640	**7468**	21624	22057	8205	7555
(1,2)	200	**4542**	4551	6815	6843	4612	4594	538	1908	**378**	2196	567	796	7497	**7303**	21278	21691	8081	7442
(1,2)	250	**4491**	4501	6763	6793	4576	4544	406	1797	**307**	1830	475	666	7348	**7151**	20946	21410	7967	7330
(1,2)	300	**4441**	4451	6710	6743	4526	4494	328	1695	**257**	1536	408	573	7202	**7042**	20621	21152	7873	7224
(1,2)	350	**4390**	4401	6658	6693	4476	4444	256	1509	**207**	1257	345	505	7083	**6907**	20328	20851	7782	7152
(1,2)	400	**4340**	4351	6608	6642	44226	4394	189	1331	**157**	1018	303	455	6991	**6815**	20031	20531	7682	7076
(1,2)	50	**4702**	5709	6975	8346	4747	5686	1672	1844	**1020**	1185	1202	1282	**8055**	9522	22324	26406	8542	10050
(2,4)	50	4690	4694	6071	6405	4687	**4685**	**109**	218	1271	1503	137	222	5766	**5243**	22547	22793	6107	5426
(4,6)	50	4689	4690	5515	5818	**4686**	4688	**22**	35	337	261	76	95	5396	**4624**	20826	21201	5363	4645
(5,10)	50	4689	4693	5323	5537	**4684**	4685	**17**	25	275	165	30	34	5381	**4574**	19405	19872	5349	4620

Results.

- For Test 1, the outcome by REPL1 show significant errors on Wiki and Cit networks, on which Greedy and REPL2 produce similar better results. While REPL2-Max stand outs for its best output on the Bitcoin network. This may be due to there is a large number of "leaves" that locate in the edge of Bitcoin network, which implies that covering them with priority can reduce the size of a broker team.
- For Test 2, it is observed that for Wiki and Cit, the size of D_1 decreases at a much slower speed as the values of ρ_1 and ρ_2 grow. This is because there are a number of isolated components with lots of zero-indegree nodes in Wiki and Cit. There, We have to assign each of such zero-indegree nodes in either D_1 and D_2 even though ρ_1 and ρ_2 are large.