A Centrality Maximization Approach for Link Recommendation

Qi Zhang
School of Information Technology and Engineering
Guangzhou College of Commerce
Guangzhou, China
Email: qzhang [AT] gcc.edu.cn

Hao Zhong*
School of Computer Science
South China Normal University
Guangzhou, China
**Corresponding author's email: scnuzhonghao [AT] foxmail.com*

Abstract—**In social networks, the goal of link recommendation is to recommend links for nodes and add them to the network, thereby satisfying the potential link interests of the nodes. The centrality of nodes in social networks typically quantifies the importance of nodes in the network. Some nodes may desire to increase their centrality by adding links. First, a multi-community centrality measurement method is proposed, and based on harmonic centrality, a hybrid centrality measurement method is introduced. Next, the link recommendation problem is regarded as a problem of maximizing node hybrid centrality, which can be formally modeled as a submodular function maximization problem. A greedy algorithm with performance guarantees can be directly applied to select the best links. Compared to existing link prediction and link recommendation algorithms, our algorithm recommends links that better improve the hybrid centrality of users.**

Keywords-social networks; link recommendation; node centrality; submodular function maximization

I. Introduction

Link recommendation is one of the many recommendation tasks in social networks, primarily targeting users' potential connection interests. For users, effective link recommendation can significantly enhance their influence or engagement within the social network. For the entire social network, adding appropriate links can improve the network's integrity and scalability, forming a more robust and scalable network. Link prediction is a widely used technique for recommending connections in social networks. At the same time, link prediction is also a hot topic in social network research, aiming to accurately predict the likelihood of a connection between two nodes in a given network. Currently, link prediction algorithms are mainly based on the network's topological structure [1,2], node attributes [3,4], or interactions between nodes [5,6].

Different link interests of nodes can provide diversified perspectives for link recommendation. In this paper, it is assumed that each node may have two different link interests. The first is to add more links to the node, making the shortest path between it and other nodes as short as possible, thus increasing its harmonic centrality [7]. The second is to add more links to the node, enabling it to connect to more communities in the network. These two link interests reflect the node's desire to integrate into the network more quickly and effectively and enhance its participation in the network. A user can either be an existing node in the network or a node that has not yet been established.

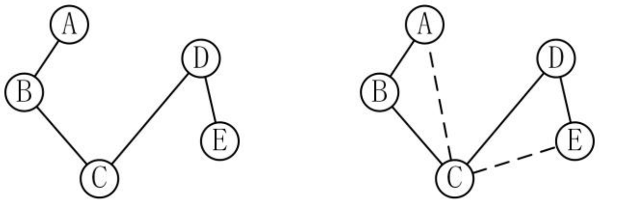
The link interests of users mentioned above give rise to a link recommendation problem, which is to increase the centrality measure of a node by adding links to it. For any given node in the network, the main reasons it seeks to enhance its centrality in the network may include the following three aspects:

(*i*) Node centrality measures the importance of a node in a social network, reflecting its role in information dissemination or connectivity. Central nodes are typically key participants in the network, influencing its overall functionality. For example, nodes with high degree centrality (connected to many other nodes) play a crucial role in information diffusion, while nodes with high betweenness centrality serve as bridges between different parts of the network.

(*ii*) Network integrity refers to the robustness of a network against failures or attacks. Studies show that the removal of central nodes can significantly disrupt network structure, especially in scale-free networks, which are resilient to random failures but vulnerable to targeted attacks on central nodes. For instance, removing nodes with high betweenness centrality may fragment the network into isolated components, reducing its overall connectivity.

(*iii*) Network scalability refers to a network's ability to maintain performance as it grows. Central nodes help sustain a small network diameter by connecting newly added nodes, ensuring efficient information flow. However, if a central node becomes overloaded with connections, it may turn into a bottleneck, limiting network expansion. In scale-free networks, central hubs are particularly important for maintaining small-world properties, supporting network scalability.

For example, consider a case where edges are added to maximize a node's degree centrality. In Figure 1(a), suppose we add two edges, C-A and C-E, to node C, directly connecting it to A, B, D, and E. As a result, node C's degree increases from 2 to 4. In a five-node graph, this is the maximum possible degree $n-1$, where n is the total number of nodes).

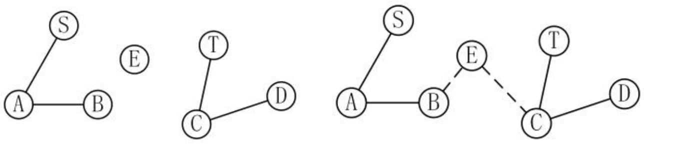


(a) A simple undirected graph (b) Adding two edges to node C

Fig. 1 An example of increasing a node's degree centrality by adding edges

This directly maximizes its degree centrality since degree centrality is simply determined by the number of edges a node has. The optimization strategy for degree centrality is intuitive and straightforward—adding more edges always increases degree centrality. In social networks, this corresponds to making a node a more directly connected "hub," such as gaining more followers or friends on social media platforms.

For example, consider a case where edges are added to maximize a node's betweenness centrality. In Figure 2(a), suppose we add edges E-B (connecting E to B in group 1) and E-C (connecting E to C in group 2), making E a bridge between the two groups. In the updated graph, any path from group 1 to group 2 (such as S-A-B-E-C-T or A-B-E-C-D) must pass through E. Since E lies on all shortest paths between the two groups, its betweenness centrality increases significantly. For instance, in the shortest path from S to T, E becomes an essential intermediary node.



(a) A simple undirected graph (b) Adding two edges to node E

Fig. 2 An example of increasing a node's betweenness centrality by adding edges

Optimizing betweenness centrality involves positioning a node as a key connector between different parts of the network. In social networks, this corresponds to making a node an "information bridge," such as an opinion leader linking two communities.

Existing methods for adding links to a graph to improve node centrality have shortcomings in mathematical analysis [8-11]. In this paper, by considering more structural information of the graph and recommending the links that should be added to the node, the process of adding nodes can be formalized as an increasing function, and the maximum value of the function can provide performance guarantees.

The aim of this study is to propose a method for maximizing centrality for link recommendation. The main contributions are as follows:

(*i*) propose a multi-community centrality measure for nodes to capture their potential interest in linking to multiple network communities;

(*ii*) integrate multi-community centrality and harmonic centrality to propose a hybrid centrality measure for nodes to comprehensively capture users' link interests;

(*iii*) model the link recommendation problem as a submodular function maximization problem based on the node's hybrid centrality measure and use a greedy algorithm to achieve an optimal link selection with an approximation ratio of $1-1/e$.

II. RELATED WORK

Suppose U is a nonempty ground set. A real set function f:$2^U \to R$ is submodular [12] if for any A, $B \subseteq U$,

$$f(A)+f(B)\geq f(A\cup B)+f(A\cap B)$$

The marginal value of an element $u \in U$ with respect to $A \subseteq U$ is defined by

$$\Delta_u f(A)=f(A\cup\{u\})-f(A)$$

Equivalently, f is submodular if it satisfies the law of diminishing marginal returns: for all $A \subseteq B \subseteq U$ and $x \notin B$,

$$\Delta_u f(A)\geq \Delta_u f(B)$$

In words, the marginal value of x diminishes as the context in which it is considered grows from A to B.

The submodular maximization problem with cardinality constraint is a critical submodular optimization problem. Given a nonempty ground set U and a monotone nonnegative submodular function f:$2^U \to R$, maximization of submodular function $f(S)$ with cardinality constraint is requires finding a subset $S \subseteq U$ with $|S|=k$. More formally,

$$\max_{S\subseteq U}\{f(S):|S|=k\}$$

The submodular maximization problem with cardinality constraint has proven to be highly effective in addressing practical issues like text abstract extraction [13], influence maximization [14], and welfare maximization [15], among others. In the domain of link recommendations, the submodular function maximization problem with cardinality constraint has also emerged as a highly effective modeling technique. Specifically, submodular function maximization problem with cardinality constraint has been employed to recommend links for nodes in order to enhance their centrality [10][16]. These existing works are aimed solely at increasing the centrality of nodes, without considering the potential interest of nodes to connect with more communities in the graph.

Submodular function maximization problem with cardinality constraint is NP-hard, which means that it is unlikely to be solved precisely in polynomial time unless $P = NP$. The classical greedy algorithm (as showed in Algorithm 1) is commonly employed for solving submodular function maximization problem with cardinality constraint. Moreover, Algorithm 1 has been proven to be an ($1-1/e$)-approximation algorithm [17].

Algorithm 1：CGASFM: Classical Greedy Algorithm for Submodular Function Maximization Problem

REQUIRE: A nonempty ground set U and a monotone nonnegative submodular function $f: 2^U \to R^+$

ENSURE: The function value $f(S)$ with $S \subseteq U$ and $|S|=k$.

1. $S \leftarrow \varnothing$
2. While $f(S) < f(D)$ do
3. select $u \in U - S$ with maximum $f(S \cup \{u\})$;
4. $S \leftarrow S \cup \{u\}$
5. Output $f(S)$

To accelerate the solution of the submodular function maximization problem with cardinality constraint without sacrificing performance guarantee, the stochastic greedy algorithm (as showed in Algorithm 2) is also proposed to solve submodular function maximization problem with cardinality constraint. Similarly, Algorithm 2 has been proven to be an ($1-1/e-\varepsilon$)-approximation algorithm[18], where ε is a parameter employed for determining the size of the sampling set.

Algorithm 2：SGASFM: Stochastic Greedy Algorithm for Submodular Function Maximization Problem

REQUIRE: A nonempty ground set U, a monotone nonnegative submodular function $f: 2^U \to R^+$ and a parameter ε

ENSURE: The function value $f(S)$ with $S \subseteq U$ and $|S|=k$.

1. $S \leftarrow \varnothing$
2. While $f(S) < f(D)$ do
3. sample a set $D \subseteq U - S$ with cardinality $\frac{|U|}{k}\log\frac{1}{\varepsilon}$ randomly
4. select $u \in D$ with maximum $f(S \cup \{u\})$ \$;
5. $S \leftarrow S \cup \{u\}$
6. Output $f(S)$

III. PROBLEM FORMULATION

Given a simple undirected graph (denoted as graph) $G=<V, E>$, where V is the set of nodes and E is the set of edges. The order of the graph G is $n=|V|$, and the size is $m=|E|$. Without loss of generality, consider recommending links for any node $v\in V$, and let $d(G, u)$ represent the shortest path length between v and u in graph G. Specifically, the following two cases need to be explained: (*i*) if there is no path connecting v and u, then $d(G, u)=\infty$ and $1/d(G, u)=0$; (*ii*) $d(G, v)=0$ and $1/d(G, v)=0$ represent the shortest path length from any node to itself being 0. Let $N(G)=\{u|(u, v)\in E\}$ be the set of all neighbor nodes of v in graph G, and let $T(G)=\{(u, v)|u \notin N(G)\}$ be the set of candidate links that can be added to graph G.

A. Multi-community centrality

Assume $C=\{C_1, C_2, \ldots, C_p\}$ is the community set of a given graph G. Generally, the community information of any graph can be obtained through existing algorithms [19-22]. Then, $N(G)\cap C_i$ represents the set of neighbor nodes of node v that belong to the i-th community C_i, where $1\leqslant i \leqslant p$. To formalize the multi-community reward of all links of node v in graph G, we use $g(u)$ to denote the relationship strength between node v and node u in graph G. This relationship strength can be measured using indicators such as their distance, similarity, etc. The multi-community centrality of node v in graph G can be defined as follows

$$l(G)=\sum_{i=1}^{p}\sqrt{\sum_{j\in N(G)\cap C_i} g(u)}\,. \tag{1}$$

To maximize the multi-community centrality of node v by adding links in graph G, it can be modeled as the maximization of a potential function $l(G)$, that is,

$$\max_{S\subseteq T(G),|S|=k} l(G_S)\,, \tag{2}$$

where $G_S=<V, E\cup S>$.

B. Harmonic Centrality

Harmonic centrality [7] is a measure used in graph theory to assess node centrality. It is calculated by summing the reciprocal of the shortest path distances between a node and all other nodes in the network. For any node $v\in V$, the harmonic centrality $h(G)$ of node v in graph G is defined as

$$h(G)=\sum_{u\in V}\frac{1}{d(G,u)}\,. \tag{3}$$

To maximize the harmonic centrality of node v by adding links in graph G, it can be modeled as the maximization of a potential function $h(G)$, that is,

$$\max_{S\subseteq T(G),|S|=k} h(G_S)\,, \tag{4}$$

where $G_S=<V, E\cup S>$.

C. Hybrid Centrality

To comprehensively consider the two potential functions $l(G)$ and $h(G)$ mentioned above, we propose a node centrality measurement method called hybrid centrality for any node $v\in V$. Its calculation is as

$$f(G)=\alpha\times l(G)+(1-\alpha)\times h(G)\,, \tag{5}$$

where $\alpha\in(0,1)$ is a parameter used to balance the computed values of $l(G)$ and $h(G)$. This is because two different situations need to be considered: (1) in different graphs, the magnitudes of $l(G)$ and $h(G)$ may differ, and ideally, they should be unified to the same scale; (2) different nodes may have different link preferences, meaning that some nodes may prefer to enhance their harmonic centrality, while others may prefer to improve their multi-community centrality.

In summary, the centrality maximization method proposed in this paper for link recommendation (A Centrality Maximization Approach for Link Recommendation, CMALR for short) can be described as follows: recommend k links for any node v such that the hybrid centrality of node v is maximized. This can be formally expressed as

$$\max_{S\subseteq T(G),|S|=k} f(G_S)\,, \tag{6}$$

where $G_S=<V, E\cup S>$.

IV. Greedy Approximation Algorithm and Complexity Analysis

We consider using a greedy approximation algorithm to identify the most suitable edges for node recommendations, aiming to enhance the node's mixed centrality. This approach intuitively reflects changes in the node's mixed centrality. In particular, the greedy approximation algorithm provides an edge set with performance guarantees, ensuring a lower bound on the increase in the node's mixed centrality. This is the motivation behind our proposed method.

Based on the potential function *f*(*G*) defined for any node *v* above, this paper considers using Algorithm 1 and Algorithm 2 to solve CMALR. To prove that Algorithms 1 and 2 are approximation algorithms, we need to show that the potential function *f*(*G*) satisfies normalized and non-decreasing, as well as submodularity. This can be done by proving that both *l*(*G*) and *h*(*G*) satisfy normalized, non-decreasing and submodularity.

Lemma 1 For any subset $S\subseteq T(G)$, potential function $l(G_S)$ satisfies:

(*i*) $l(G_S)\geqslant 0$;

(*ii*) $l(G_B)\geqslant l(G_A)$ for any subset $A\subseteq B\subseteq T(G)$

(*iii*) $\Delta_{(v,u)}l(G_A)\ge\Delta_{(v,u)}l(G_B)$ for any subset $A\subseteq B\subseteq T(G)$ and any $(v,u)\in T(G)-B$.

Proof. (*i*) By definition, $g(u)$ represents the relationship strength between node *v* and node *u* in graph *G*. Typically, this measure ensures that $g(u)\geqslant 0$, and it is evident that $l(G_S)\geqslant 0$.

(*ii*) For any set $A\subseteq B\subseteq T(G)$ and any $1\leqslant i\leqslant p$, it can be seen that $N(G_A)\cap C_i\subseteq N(G_B)\cap C_i$, and thus $l(G_B)\geqslant l(G_A)$ holds.

(*iii*)Assume that node *u* belongs to the *i*-th community C_i, then

$$\begin{aligned}&\Delta_{(v,u)}l(G_A)-\Delta_{(v,u)}l(G_B)\\&=(\sqrt{\sum_{j\in N(G_A)\cup\{u\}\cap C_i}g(j)}-\sqrt{\sum_{j\in N(G_A)\cap C_i}g(j)})\\&-\sqrt{\sum_{j\in N(G_B)\cup\{u\}\cap C_i}g(j)}-\sqrt{\sum_{j\in N(G_B)\cap C_i}g(j)}\\&=\frac{g(u)}{\sqrt{\sum_{j\in N(G_A)\cup\{u\}\cap C_i}g(j)}+\sqrt{\sum_{j\in N(G_A)\cap C_i}g(j)}}\\&-\frac{g(u)}{\sqrt{\sum_{j\in N(G_B)\cup\{u\}\cap C_i}g(j)}+\sqrt{\sum_{j\in N(G_B)\cap C_i}g(j)}}\end{aligned}\,, \tag{7}$$

Due to

$$\sqrt{\sum_{j\in N(G_A)\cup\{u\}\cap C_i}g(j)}\le\sqrt{\sum_{j\in N(G_B)\cup\{u\}\cap C_i}g(j)} \tag{8}$$

and

$$\sqrt{\sum_{j\in N(G_A)\cap C_i}g(j)}\le\sum_{j\in N(G_B)\cap C_i}g(j) \tag{9}$$

Then, it is easy to deduce that for any $A\subseteq B\subseteq T(G)$ and $(v,u)\in T(G)-B$, the condition $\Delta_{(v,u)}l(G_A)\ge\Delta_{(v,u)}l(G_B)$ holds.

□

Lemma 2 For any subset $S\subseteq T(G)$, potential function $h(G_S)$ satisfies:

(*i*) $h(G_S)\geqslant 0$;

(*ii*) $h(G_B)\geqslant h(G_A)$ for any subset $A\subseteq B\subseteq T(G)$

(*iii*) $\Delta_{(v,t)}h(G_A)\ge\Delta_{(v,t)}h(G_B)$ for any subset $A\subseteq B\subseteq T(G)$ and any $(v,u)\in T(G)-B$.

Proof. (*i*)Since for any $u\in V$, $1/d(G_S, u)\geqslant 0$, it is evident that $h(G_S)\geqslant 0$.

(*ii*) For any set $A\subseteq B\subseteq T(G)$ and any $u\in V$, it can be seen that $d(G_A, u)\geqslant d(G_B, u)$, and thus $h(G_B)\geqslant h(G_A)$ holds.

(*iii*)For any two nodes *t* and *u*, let *dis*(*t*,*u*) represent the shortest path length between them in graph *G*. Then

$$h(G)=\sum_{u\in V}\frac{1}{d(G,u)}=\sum_{u\in V}\frac{1}{1+\min_{w\in N(G)} dis(w,u)} \tag{10}$$

Then

$$\begin{aligned}&\Delta_{(v,t)}h(G_A)\\&=\sum_{u\in V}\frac{1}{1+\min\{\min_{w\in N(G_A)} dis(w,u), dis(t,u)\}}\\&-\sum_{u\in V}\frac{1}{1+\min_{w\in N(G_A)} dis(w,u)}\end{aligned} \tag{11}$$

and

$$\begin{aligned}&\Delta_{(v,t)}h(G_B)\\&=\sum_{u\in V}\frac{1}{1+\min\{\min_{w\in N(G_B)} dis(w,u), dis(t,u)\}}\\&-\sum_{u\in V}\frac{1}{1+\min_{w\in N(G_B)} dis(w,u)}\end{aligned} \tag{12}$$

Next, we will discuss the following three cases:

(1) $dis(t,u)\geq \min_{w\in N(G_A)} dis(w,u)$, in which case $\Delta_{(v,t)}h(G_A)=\Delta_{(v,t)}h(G_B)=0$.

(2) $dis(t,u)<\min_{w\in N(G_A)} dis(w,u)$ and $dis(t,u)\geq \min_{w\in N(G_B)} dis(w,u)$, in which case $\Delta_{(v,t)}h(G_A)>0=\Delta_{(v,t)}h(G_B)$.

(3) $dis(t,u)<\min_{w\in N(G_A)} dis(w,u)$ and $dis(t,u)<\min_{w\in N(G_B)} dis(w,u)$, in which case it is easy to see that $\Delta_{(v,t)}h(G_A)\geq \Delta_{(v,t)}h(G_B)$ because $\min_{w\in N(G_A)} dis(w,u)\geq \min_{w\in N(G_B)} dis(w,u)$.

In summary, we can conclude that for any $A\subseteq B\subseteq T(G)$ and $(v,t)\in T(G)-B$, $\Delta_{(v,t)}h(G_A)\geq \Delta_{(v,t)}h(G_B)$ holds.

□

From Lemma 1 and Lemma 2, it can be inferred that the potential function *f* is normalized and non-decreasing, as well as submodular because it is the summation of the potential functions *l* and *h*. Therefore, Algorithms 1 or 2 can be used to find an approximate solution to the maximum of the potential function *f*, with performance guarantees. That is, Algorithm 1 achieves an approximation ratio of $1-1/e$ for solving CMALR, while Algorithm 2 achieves an approximation ratio of $1-1/e-\varepsilon$.

As shown in Algorithm 1, let *n* and *m* be the number of nodes and edges in the input graph *G*. For any node $v\in V$, $h(G)$ can be computed in polynomial time $O(n^2)$. Additionally, the calculation of $l(G)$ for node *v* can be completed in $O(n)$ time. Therefore, the calculation of $\Delta_{(v,t)}f(G_S)$ for node *v* can be done in $O(n^2+n+m\log m)$ time. The time complexity of Algorithm 1 is $O(kn^2\log n)$, since the number of iterations is *k*. Assuming the sample size of Algorithm 2 is $t=\frac{n}{k}\log\frac{1}{\varepsilon}$, the time complexity of Algorithm 2 is $O(kt^2\log t)$.

V. Experiment

This section aims to experimentally evaluate the actual performance of typical social network instances in terms of solution quality. All algorithms used for comparison are implemented in Python and run on a computer with a 2.10GHz Intel Core i7-12700 processor and 16 GB RAM.

A. Datasets

These datasets contains 8 real social network instances from the Stanford network dataset collection. For SCHOLAT Social Network, we consider extracting a subgraph with the highest degree consisting of 150 nodes from graph. These real network instances have been widely used to validate various algorithms. Table 1 shows the detailed information of these real social networks.

Table 1 Typical Social Networks Graphs Used in Experiment

Networks	Nodes	Edges	Average degree
Adjnoun[23]	112	425	7.59
Dolphin[24]	62	159	5.13
Football[25]	115	613	10.66
Karate[26]	34	78	4.59
Lesmis[27]	77	254	6.60
Polbooks[28]	105	441	8.40
Celegansneural[29]	297	2148	14.46
SCHOLAT[30]	150	7808	104.11

B. Experimental Results and Analysis

This section presents the main results of Algorithm 1 (Greedy), eight link prediction algorithms (Resource [31], Jaccard [32], Adamic [32], Preference [32], Hopcroft [33], Soundarajan [33], Within [34], Common [35]), and one link recommendation algorithm (Crescenzi [10]) for solving CMALR. The experimental results are shown in Tables 2 and 3. Based on all the algorithms mentioned above, we recommend 5 links for all nodes in V, and then compare the average values of all potential functions *f*.

Table 2 Recommended Results of All Algorithms on Four Datasets

	Adjnoun	Dolphin	Football	Karate
Adamic	149.53	83.06	101.71	48.53
Soundarajan	117.22	77.48	96.45	45.04
Jaccard	136.23	80.22	109.27	57.75
Preference	171.82	98.79	109.27	57.72
Hopcroft	134.51	76.24	94.20	44.01
Resource	148.39	83.14	101.71	48.48
Within	116.27	76.59	94.28	43.97
Common	151.51	82.57	102.21	49.53
Crescenzi	266.79	151.21	235.29	82.98
Greedy	**343.42**	**202.25**	**265.23**	**97.67**

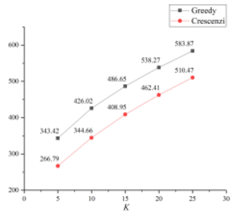
Table 3 Recommended Results of All Algorithms on Other Four Datasets

	Lesmis	Polbooks	Celegansneural	SCHOLAT
Adamic	113.64	111.09	149.51	288.59

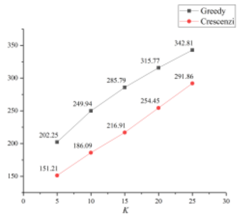
Soundarajan	102.87	109.85	149.97	269.69
Jaccard	99.22	106.82	149.59	268.35
Preference	148.10	134.02	148.98	317.19
Hopcroft	93.21	107.01	146.22	263.59
Resource	113.36	111.96	149.65	288.03
Within	93.19	108.43	145.22	263.51
Common	115.73	112.24	149.54	263.51
Crescenzi	204.09	235.59	287.79	589.70
Greedy	**272.91**	**296.43**	**293.23**	**700.48**

The results in Tables 2 and 3 support our proposed algorithm, which not only theoretically guarantees performance, but also achieves better solutions on typical real-world social network instances.

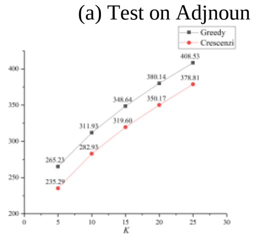
In the second experiment, we consider recommending K =5, 10, 15, 20, 25 links for each node $v \in V$. We compared the approximate results of the maximum value of the potential energy function f for Greedy and Crescenzi, as shown in Figure 3.



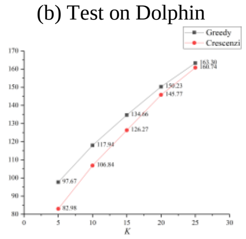
(a) Test on Adjnoun



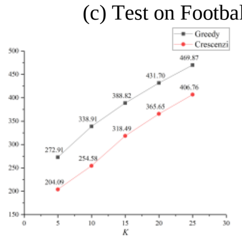
(b) Test on Dolphin



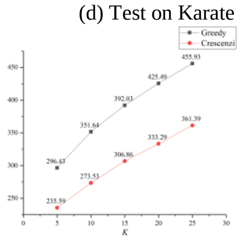
(c) Test on Football



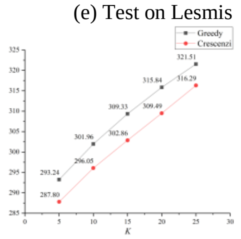
(d) Test on Karate



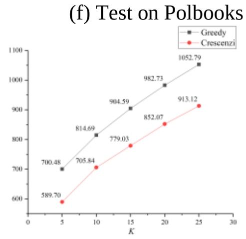
(e) Test on Lesmis



(f) Test on Polbooks



(g) Test on Celegansneural

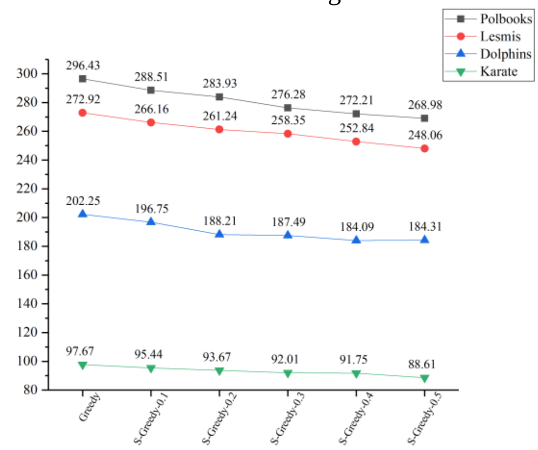


(h) Test on SCHOLAT

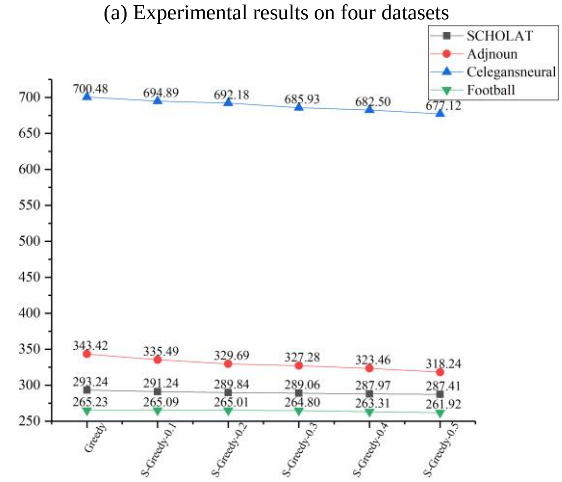
Fig. 3 The experimental results for K=5,10,15,20 and 25 links are recommended on different datasets.

The results in Figure 3 support our proof of Lemma 1 and Lemma 2, which states that the potential function f is normalized and non-decreasing relative to the set S. Meanwhile, the results in Figure 3 also indicate that Greedy can obtain better solutions by recommending different numbers of links on typical real-world social network instances.

In the final set of experiments, we consider recommending K=5 links for each node $v \in V$ based on Algorithm 1 (Greedy) and Algorithm 2 (S-Greedy-ε). Set ε to 0.1, 0.2, 0.3, 0.4, and 0.5, and the corresponding experimental results are shown in Figure 4.



(a) Experimental results on four datasets



(b) Experimental results on the other four datasets

Fig. 4 The experimental results for K=5 links are recommended on different datasets based on algorithm 1 and algorithm 2.

The results in Figure 4 indicate that the performance guarantee of Algorithm 2 decreases as ε increases. This is mainly because the random sampling range in Algorithm 2 decreases as ε increases. However, compared to Algorithm 1, Algorithm 2 has lower time complexity, so Algorithm 2 can also be prioritized for use considering time consumption conditions.

VI. CONCLUSION

In real-world social networks, centrality analysis is widely used to identify key nodes. For example, in epidemic spread studies, identifying high-centrality "superspreader" nodes helps control disease transmission. On social media platforms, central nodes (such as opinion leaders) play a crucial role in information dissemination, but their removal may impact network connectivity and scalability. Node centrality is crucial in social networks, as its impact on network integrity and scalability reflects the complexity of network structures. Future research could further explore how the dynamic changes in centrality distribution affect the long-term robustness and scalability of networks, especially in large-scale dynamic networks.

This paper comprehensively considers adding links to nodes in order to increase both multi-community centrality and harmonic centrality, and proposes a centrality maximization approach for link recommendation. We model the link recommendation problem as a submodular function maximization problem by designing a submodular potential function. Then, we directly present classical greedy algorithms and random greedy algorithms with performance guarantees to select the optimal links. Compared with 8 link prediction algorithms and 1 link recommendation algorithm, Algorithm 1 can better improve the centrality of users, enabling them to connect to more communities within the social network.

In future work, we aim to consider more link interests within the submodular potential function. This way, we can recommend links to nodes through greedy algorithms while ensuring performance guarantees.

ACKNOWLEDGMENT

This research was funded by the National Key Research and Development Program of China (Research and Demonstration Application of Key Technologies for Personalized Learning Driven by Educational Big Data) under Grant 2023YFC3341200, the National Natural Science Foundation of China under Grant 62377015, the Collaborative Innovation Center for Intelligent Educational Technology of Guangzhou under Grant 2023B04J0002, Tertiary Education Scientific research project of Guangzhou Municipal Education Bureau 2024312300, the National Natural Science Foundation of China under Grant 62407016, and the Research Cultivation Fund for The Youth Teachers of South China Normal University under Grant 23KJ29.

REFERENCES

[1] Mishra S, Singh S S, Kumar A, et al. HOPLP－MUL: link prediction in multiplex networks based on higher order paths and layer fusion[J]. Applied Intelligence, 2023, 53(3): 3415-3443.

[2] Zhou M, Jin H, Wu Q, et al. Betweenness centrality-based community adaptive network representation for link prediction[J]. Applied Intelligence, 2022, 52(4): 3545-3558.

[3] Le T, Le N, Le B. Knowledge graph embedding by relational rotation and complex convolution for link prediction[J]. Expert Systems with Applications, 2023, 214: 119122.

[4] Zhang J, Huang J, Gao J, et al. Knowledge graph embedding by logical-default attention graph convolution neural network for link prediction[J]. Information Sciences, 2022, 593: 201-215.

[5] Zhao A, Yu Y. Context aware sentiment link prediction in heterogeneous social network[J]. Cognitive Computation, 2022, 14(1): 300-309.

[6] Liu Q, Chen Y, Zhang G, et al. A novel functional network based on three-way decision for link prediction in signed social networks[J]. Cognitive Computation, 2022, 14(6): 1942-1954.

[7] Boldi P, Vigna S. Axioms for centrality[J]. Internet Mathematics, 2014, 10(3-4): 222-262.

[8] Papagelis M, Bonchi F, Gionis A. Suggesting ghost edges for a smaller world[C]//Proceedings of the 20th ACM international conference on Information and knowledge management. 2011: 2305-2308.

[9] Parotsidis N, Pitoura E, Tsaparas P. Selecting shortcuts for a smaller world[C]//Proceedings of the 2015 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2015: 28-36.

[10] Crescenzi P, d′Angelo G, Severini L, et al. Greedily improving our own centrality in a network[C]//International symposium on experimental algorithms. Cham: Springer International Publishing, 2015: 43-55.

[11] Ishakian V, Erdös D, Terzi E, et al. A framework for the evaluation and management of network centrality[C]//Proceedings of the 2012 SIAM International Conference on Data Mining. Society for Industrial and Applied Mathematics, 2012: 427-438.

[12] Chen W, Zhong H, Wu L, et al. A general greedy approximation algorithm for finding minimum positive influence dominating sets in social networks[J]. Journal of Combinatorial Optimization, 2022, 44(1): 1-20.

[13] Lin H, Bilmes J. A class of submodular functions for document summarization[C]//Proceedings of the 49th annual meeting of the association for computational linguistics: human language technologies. 2011: 510-520.

[14] Yang W, Zhang Y, Du D Z. Influence maximization problem: properties and algorithms[J]. Journal of Combinatorial Optimization, 2020, 40(4): 907-928.

[15] Chakrabarty D, Goel G. On the approximability of budgeted allocations and improved lower bounds for submodular welfare maximization and GAP[J]. SIAM Journal on Computing, 2010, 39(6): 2189-2211.

[16] Parotsidis N, Pitoura E, Tsaparas P. Centrality-aware link recommendations[C]//Proceedings of the ninth ACM international conference on web search and data mining. 2016: 503-512.

[17] Wolsey L A. An analysis of the greedy algorithm for the submodular set covering problem[J]. Combinatorica, 1982, 2(4): 385-393.

[18] Mirzasoleiman B, Badanidiyuru A, Karbasi A, et al. Lazier than lazy greedy[C]//Proceedings of the AAAI Conference on Artificial Intelligence. 2015: 1812–1818.

[19] Palla G, Derényi I, Farkas I, et al. Uncovering the overlapping community structure of complex networks in nature and society[J]. nature, 2005, 435(7043): 814-818.

[20] Reichardt J, Bornholdt S. Statistical mechanics of community detection[J]. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 2006, 74(1): 016110.

[21] Cordasco G, Gargano L. Community detection via semi-synchronous label propagation algorithms[C]//2010 IEEE international workshop on: business applications of social network analysis (BASNA). IEEE, 2010: 1-8.

[22] Traag V A, Waltman L, Van Eck N J. From Louvain to Leiden: guaranteeing well-connected communities[J]. Scientific reports, 2019, 9(1): 1-12.

[23] Newman M E J. Finding community structure in networks using the eigenvectors of matrices[J]. Physical Review E—Statistical, Nonlinear, and Soft Matter Physics, 2006, 74(3): 036104.

[24] Lusseau D, Schneider K, Boisseau O J, et al. The bottlenose dolphin community of doubtful sound features a large proportion of long-lasting associations: can geographic isolation explain this unique trait?[J]. Behavioral Ecology and Sociobiology, 2003, 54: 396-405.

[25] Girvan M, Newman M E J. Community structure in social and biological networks[J]. Proceedings of the national academy of sciences, 2002, 99(12): 7821-7826.

[26] Zachary W W. An information flow model for conflict and fission in small groups[J]. Journal of anthropological research, 1977, 33(4): 452-473.

[27] Liu B, Xu S, Li T, et al. Quantifying the effects of topology and weight for link prediction in weighted complex networks[J]. Entropy, 2018, 20(5): 363.

[28] Ye Q, Zhu C, Li G, et al. Using node identifiers and community prior for graph-based classification[J]. Data Science and Engineering, 2018, 3(1): 68-83.

[29] Watts D J, Strogatz S H. Collective dynamics of ʻsmall-worldʼ networks[J]. nature, 1998, 393(6684): 440-442.

[30] Xu Q, Qiu L, Lin R, et al. An improved community detection algorithm via fusing topology and attribute information[C]//2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD). IEEE, 2021: 1069-1074.

[31] Zhou T, Lü L, Zhang Y C. Predicting missing links via local information[J]. The European Physical Journal B, 2009, 71: 623-630.

[32] Liben-Nowell D, Kleinberg J. The link prediction problem for social networks[C]//Proceedings of the twelfth international conference on Information and knowledge management. 2003: 556-559.

[33] Soundarajan S, Hopcroft J. Using community information to improve the precision of link prediction methods[C]//Proceedings of the 21st international conference on world wide web. 2012: 607-608.

[34] Valverde-Rebaza J C, de Andrade Lopes A. Link prediction in complex networks based on cluster information[C]//Advances in Artificial Intelligence-SBIA 2012: 21th Brazilian Symposium on Artificial Intelligence, Curitiba, Brazil, October 20-25, 2012. Proceedings. Springer Berlin Heidelberg, 2012: 92-101.

[35] Ahmad I, Akhtar M U, Noor S, et al. Missing link prediction using common neighbor and centrality based parameterized algorithm[J]. Scientific reports, 2020, 10(1): 364.