LinkBlackHole*: Robust Overlapping Community Detection Using Link Embedding (Extended Abstract)*

Jungeun Kim
KAIST
je_kim@kaist.ac.kr

Sungsu Lim
Chungnam National University
sungsu@cnu.ac.kr

Jae-Gil Lee◇
KAIST
jaegil@kaist.ac.kr

Byung Suk Lee
University of Vermont
bslee@uvm.edu

***Abstract*—This paper proposes LinkBlackHole*, a novel algorithm for finding communities that are (i) overlapping in nodes and (ii) mixing (not separating clearly) in links. There has been a small body of work in each category, but this paper is the first one that addresses both. For this purpose, LinkBlackHole* incorporates the advantages of both the link-space transformation and the black hole transformation. Thorough experiments show superior quality of the communities detected by LinkBlackHole* to those detected by other state-of-the-art algorithms.**

I. Introduction

Overlapping community detection is known to be harder than disjoint community detection [1], [2]. Currently, LinkSCAN* [2], which introduced the novel concept of *link-space transformation*, has been regarded as the state-of-the-art algorithm. It transforms the original graph to a link-space graph, where each node is mapped from a link (i.e., edge) in the original graph in such a way that two nodes are adjacent if the corresponding links in the original graph share a common end node. The benefit of the transformation is that *disjoint* community detection from the link-space graph produces *overlapping* communities.

Overlapping community detection is further complicated by another issue: *mixing* of links between communities. Quantitatively, mixing is defined as the fraction of links in the network that are crossing between different communities [3]. Such links ambiguate the detection of boundaries between communities and consequently make community structures harder to detect. Lim et al. [4] resolved the mixing problem by incorporating a *geometric embedding* technique in their algorithm BlackHole, which we call the *black hole transformation* henceforth. Multiple nodes that are likely to belong to the same community are mapped to (almost) the same position (dubbed a "black hole") as a result of the embedding. Clustering is then performed on the resulting positions.

One intuitive idea for achieving robust overlapping community detection despite the mixing effect is to combine (1) the link-space transformation [2] and (2) the black hole transformation [4]. We propose this combined algorithm and call it *LinkBlackHole**. The algorithm is essentially *link* embedding that is done by BlackHole when it works on the link-space graph output by LinkSCAN* because the nodes embedded are mapped from links in the original graph.

∗ This research was supported by the MOLIT (The Ministry of Land, Infrastructure and Transport), Korea, under the national spatial information research program supervised by the KAIA (Korea Agency for Infrastructure Technology Advancement) (19NSIP-B081011-06).

◇ Jae-Gil Lee is the corresponding author.

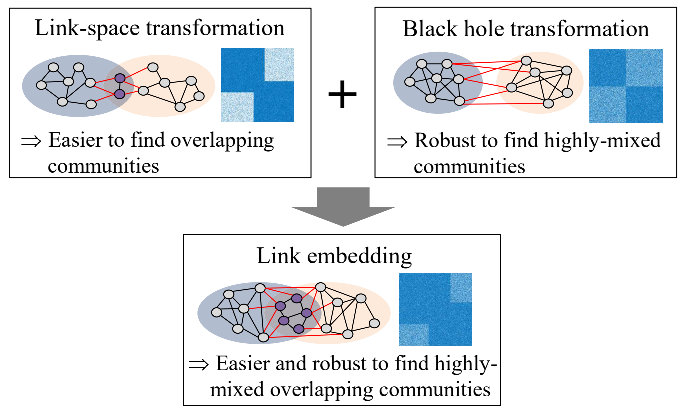

Fig. 1. The advantage of link embedding in LinkBlackHole*.

II. Algorithm Overview

Figure 1 summarizes the merit of link embedding in LinkBlackHole*, inherited from the link-space transformation of LinkSCAN* and the black hole transformation of BlackHole. Concretely, the link-space transformation enables us to accurately find *overlapping* communities in the original graph through finding *disjoint* communities in the link-space graph while preserving the original graph structure. In turn, the black hole transformation achieves robust accuracy of finding *highly-mixed* overlapping communities.

We first present LinkBlackHole, a robust overlapping community detection algorithm based on link embedding, and then its variant LinkBlackHole*, which improves the clustering efficiency of LinkBlackHole by reducing the number of links through random sampling while maintaining comparable clustering accuracy. This algorithm development structure parallels that of LinkSCAN and LinkSCAN* [2].

LinkBlackHole* has four main steps, as illustrated in Figure 2. First, the original graph $\mathcal{G}$ is converted to a link-space graph $LS(\mathcal{G})$ by the link-space transformation, and the links of $LS(\mathcal{G})$ are sampled. (Please note that this sampling is not part of LinkBlackHole.) Second, every node in the sampled link-space graph $LS'(\mathcal{G})$ is mapped to a point in a low dimensional space. Third, the positions are clustered using a conventional clustering algorithm such as DBSCAN [5], and link-space

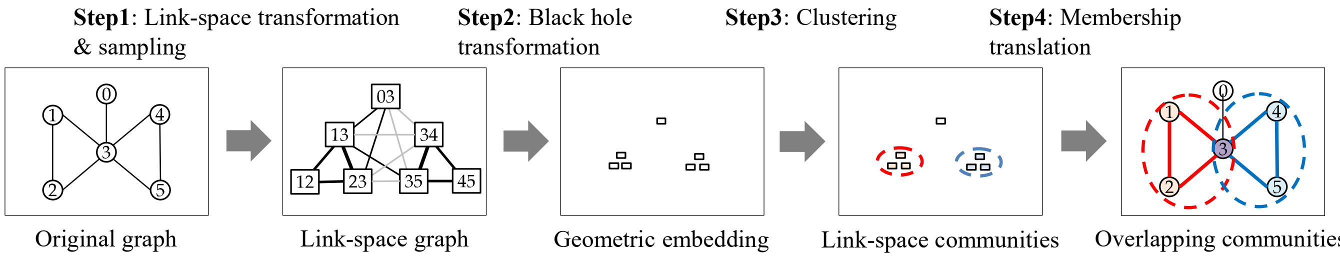

Fig. 2. The main steps of LinkBlackHole*.

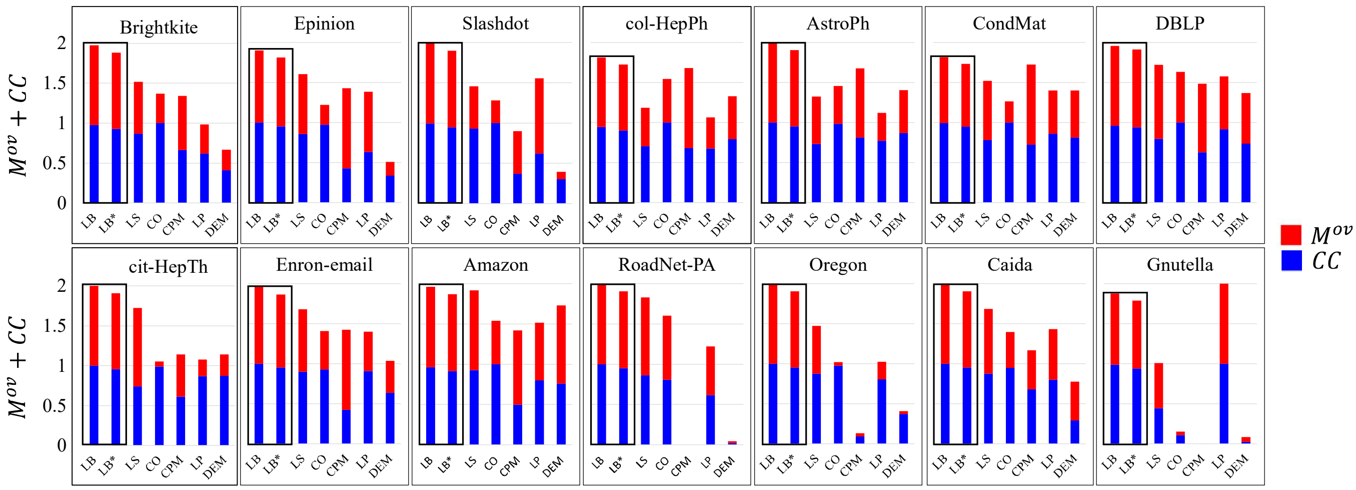

Fig. 3. Results for real-world networks.

communities in $LS'(\mathcal{G})$ are determined from the clusters. Fourth, the link membership in $LS'(\mathcal{G})$ is translated back to node membership in $\mathcal{G}$.

III. Evaluation Result

We use two metrics to quantify the quality of clusters. *Overlapping modularity* M^{ov} [6] measures, for each cluster, how densely modularized (i.e., connected inward and disconnected outward) the connected nodes are in the cluster. Since the modularity is favorable to a small community, we employ another counter-balancing metric. *Clustering coverage* CC [7] is the fraction of nodes having a cluster membership.

LinkBlackHole (LB) and LinkBlackHole* (LB*) are compared against LinkSCAN* (LS) [2], COPRA (CO) [8], CPM [9], LinkPartition (LP) [7], and DEMON (DEM) [10].

Figure 3 shows the results for 14 real-world networks. M^{ov} and CC, each normalized to the range of 0 to 1, are overlaid in the same plot so that their values can be compared between algorithms not only as a whole measure but also in separate measures. The results are impressive. LinkBlackHole* and LinkBlackHole show significantly better performance than the other algorithms for 13 out of 14 data sets. Even in the one exception for Gnutella, LinkPartition shows only *slightly* (6%) better performance than LinkBlackHole* and LinkBlackHole, which are the close second best performers.

IV. Conclusion

LinkBlackHole* combines LinkSCAN*'s ability to detect overlapping communities and BlackHole's robustness to mixing between communities, and thereby can accurately detect highly-mixed overlapping communities from large networks. Extensive evaluation demonstrated the performance merit of LinkBlackHole* compared with other algorithms that define the state of overlapping community detection.

References

[1] J. Xie, S. Kelley, and B. K. Szymanski, "Overlapping community detection in networks: the state of the art and comparative study," *ACM Computing Surveys*, vol. 45, no. 4, p. 43, 2013.

[2] S. Lim, S. Ryu, S. Kwon, K. Jung, and J.-G. Lee, "LinkSCAN*: Overlapping community detection using the link-space transformation," in *Proc. 30th IEEE Int'l Conf. on Data Engineering*, 2014, pp. 292–303.

[3] A. Lancichinetti, S. Fortunato, and J. Kertész, "Detecting the overlapping and hierarchical community structure in complex networks," *New Journal of Physics*, vol. 11, p. 033015, 2009.

[4] S. Lim, J. Kim, and J.-G. Lee, "BlackHole: Robust community detection inspired by graph drawing," in *Proc. 32nd IEEE Int'l Conf. on Data Engineering*, 2016, pp. 25–36.

[5] M. Ester, H.-P. Kriegel, J. Sander, M. Wimmer, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise," in *Proc. 2nd ACM SIGKDD Int'l Conf. on Knowledge Discovery and Data Mining*, 1996, pp. 291–316.

[6] H.-W. Shen, X.-Q. Cheng, and J.-F. Guo, "Quantifying and identifying the overlapping community structure in networks," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2009, no. 7, p. P07042, 2009.

[7] Y.-Y. Ahn, J. P. Bagrow, and S. Lehmann, "Link communities reveal multiscale complexity in networks," *Nature*, vol. 466, no. 7307, pp. 761–764, 2010.

[8] S. Gregory, "Finding overlapping communities in networks by label propagation," *New Journal of Physics*, vol. 12, p. 103018, 2010.

[9] J. M. Kumpula, M. Kivelä, K. Kaski, and J. Saramäki, "Sequential algorithm for fast clique percolation," *Physical Review E*, vol. 78, no. 2, p. 026109, 2008.

[10] M. Coscia, G. Rossetti, F. Giannotti, and D. Pedreschi, "Uncovering hierarchical and overlapping communities with a local-first approach," *ACM Trans. on Knowledge Discovery from Data*, vol. 9, no. 1, p. 6, 2014.