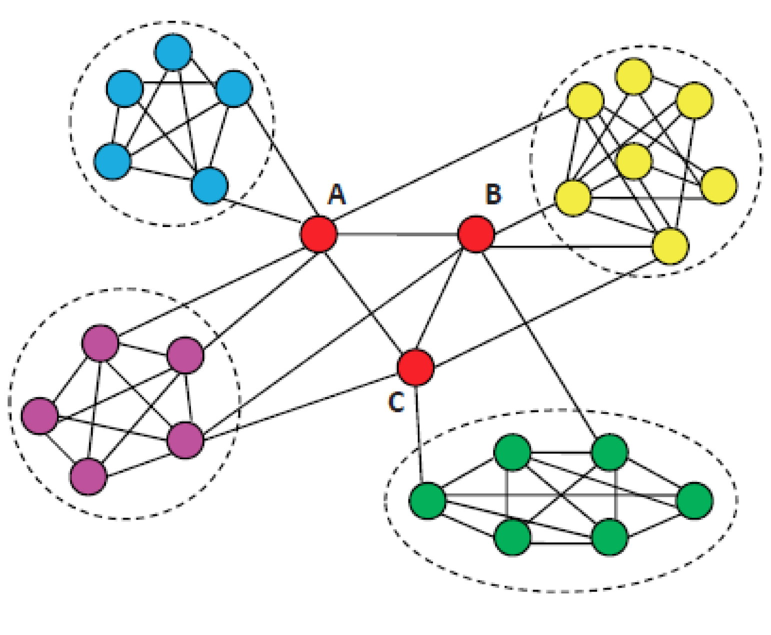
An Embedding Approach to Anomaly Detection

Renjun Hu[1], Charu Aggarwal[2], Shuai Ma[1,*], Jinpeng Huai[1]
[1]SKLSDE Lab, Beihang University, China [2]IBM T. J. Watson Research Center, USA
{hurenjun, mashuai, huaijp}@buaa.edu.cn charu@us.ibm.com

Introduction

In this paper, we are to discover *structural inconsistencies, i.e., nodes that connect to a number of diverse influential communities in the network* (Fig. 1).



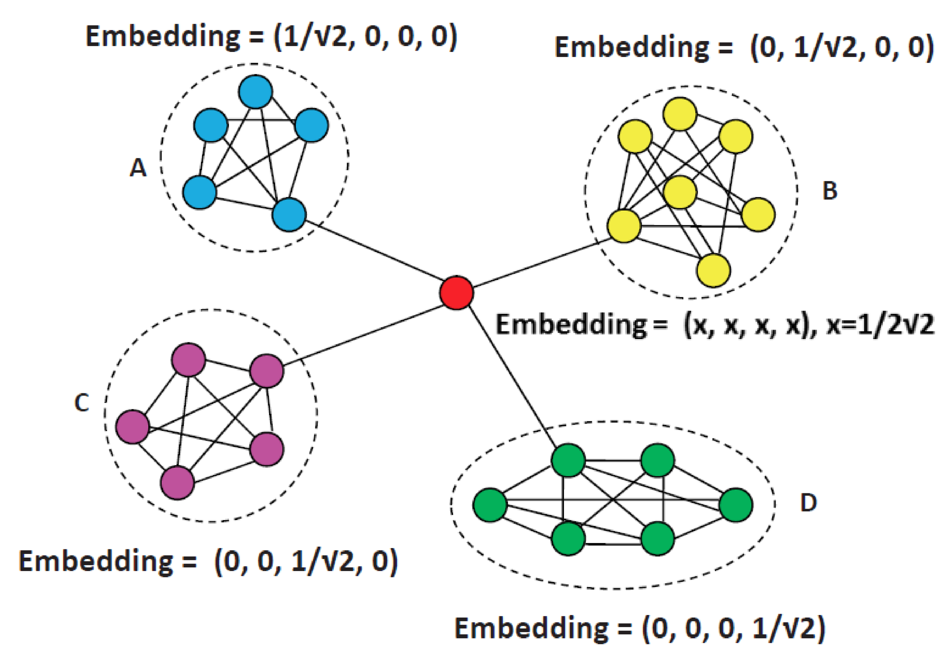


Figure 1: anomalous (red) nodes **Figure 2:** nodes in embedding

Relationship with structural hole brokers. In Burt's structural hole theory, an individual (broker) who acts as a mediator between two or more groups of people (*e.g.*, A-C in Fig. 1) would gain important social capital such as novel ideas [1]. In this sense, structural inconsistencies also provide a formal definition for structural hole brokers.

Impacts of anomalies. Since the anomalous nodes connect to diverse regions in the network, the incident links violates the notion of *homophily* [2], which assumes that linked nodes have similar properties. Because of this inconsistency in the link structures, the presence of such anomalies may:

- have a substantial impact on network structure, *e.g.*, nodes from four groups tend to form one large group in Fig. 1;
- prevent effective application of many network mining algorithms, *e.g.*, hard to achieve meaningful clusters.

Graph Embedding

To detect structural inconsistencies, we first use graph embedding to associate each node with a multidimensional position. In the embedding model, each dimension corresponds to a clustered region in the network.

Given an undirected graph $G=(V,E)$, associate each node i with a d-dimensional vector X_i, which represents the correlation between node i and the d communities (Fig. 2). The *goal* in this embedding is to preserve linkage structure of the network. Finally, the embedding is determined by *minimizing the objective function O*:

$$O = \sum_{(i,j)\in E} \|X_i - X_j\|^2 + \alpha \cdot \sum_{(i,j)\notin E} \left(1 - \|X_i - X_j\|\right)^2,$$

where α is a balancing factor, which regulates the importance of the two components in O.

A Quantitative Measure of Anomaly

After deriving the embedding, anomalous nodes are determined using the embedding together with a quantitative measure.

We first define *NB(i)* to evaluate the correlation of node i with the d communities (instead of using X_i alone):

$$NB(i) = \left(y_i^1, \ldots, y_i^d\right) = \sum_{(i,j)\in E} \left(1 - \|X_i - X_j\|\right) \cdot X_j$$

Given *NB(i)*, we introduce the *AScore* measure to indicate the anomalousness level of node i:

$$AScore(i) = \sum_{k=1}^{d} \frac{y_i^k}{y_i^*},\ y_i^* = \max\left\{y_i^1, \ldots, y_i^d\right\}$$

Finally, node i is detected as an anomaly if *AScore(i)>thre*.

In fact, *AScore* measure is also a quantitative measure for detecting structural hole brokers.

Algorithm Optimizations

By now, the $O(n^2)$ terms in O make our approach hard to be applied to large networks. Hence, we further use the sampling and graph partitioning, and propose a novel dimension reduction technique, to make our approach more scalable and effective for large networks.

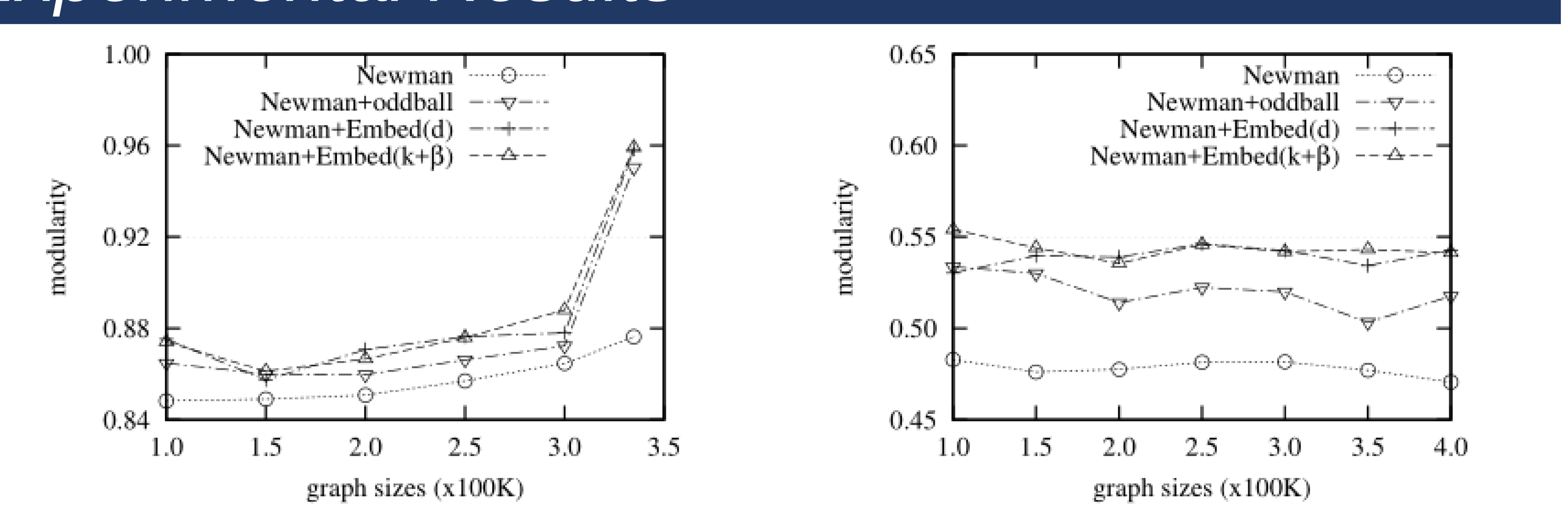
Sampling. It is very inefficient to express O as a sum of $O(n^2)$ terms. An observation here is that α is typically picked close to 0 and it is possible to approximately represent O by sampling a subset E_s of size $|E|$ for the second component:

$$O \approx \sum_{(i,j)\in E} \|X_i - X_j\|^2 + \sum_{(i,j)\in E_s} \left(1 - \|X_i - X_j\|\right)^2, E_s \subset \{(i,j) \mid (i,j) \notin E\}$$

Graph partitioning based initialization. We use a gradient descent method to optimize O, which is critically dependent on a good initialization. Thus, we incorporate graph partitioning for initialization such that densely connected nodes are initialized with similar embedding values (Fig. 2).

Dimension reduction. The number d can be large in practice, while anomalies typically connect to a limited number of communities. This motivate us to only maintain ($k+\beta$)-dimensions for embedding of each node. Numbers k and β could be much smaller, *e.g.,* 10 and 2.

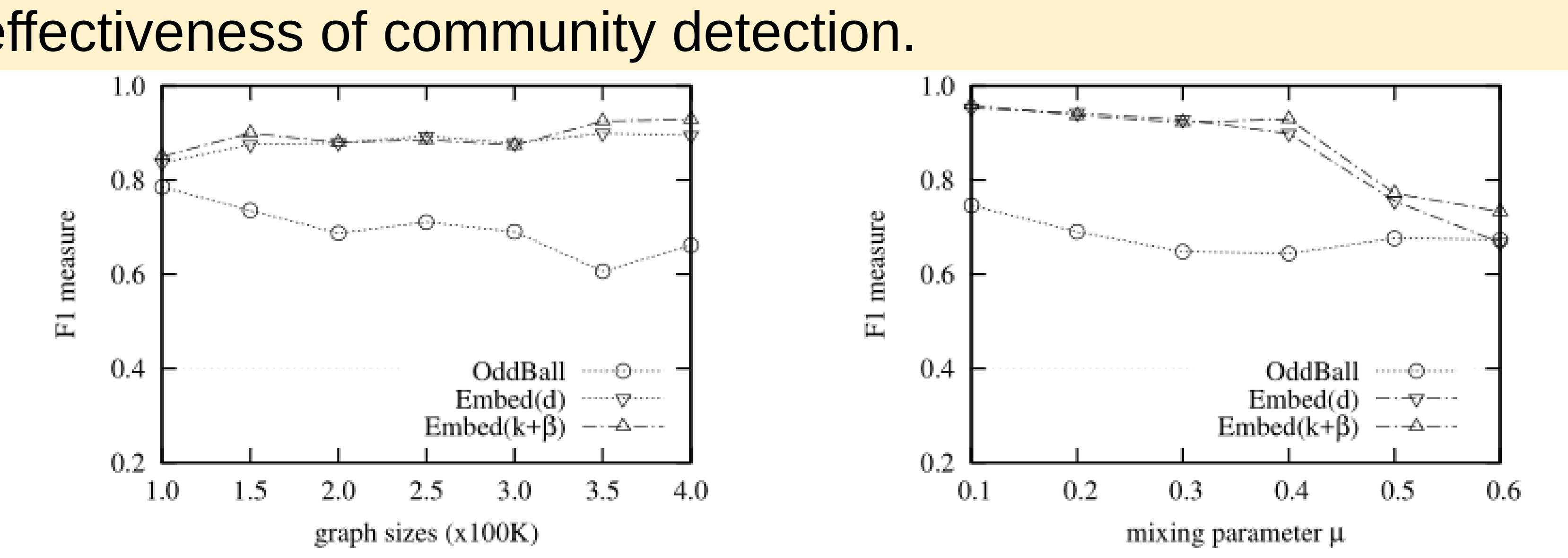
Experimental Results



(a) AMAZON dataset (b) SYNTHETIC dataset

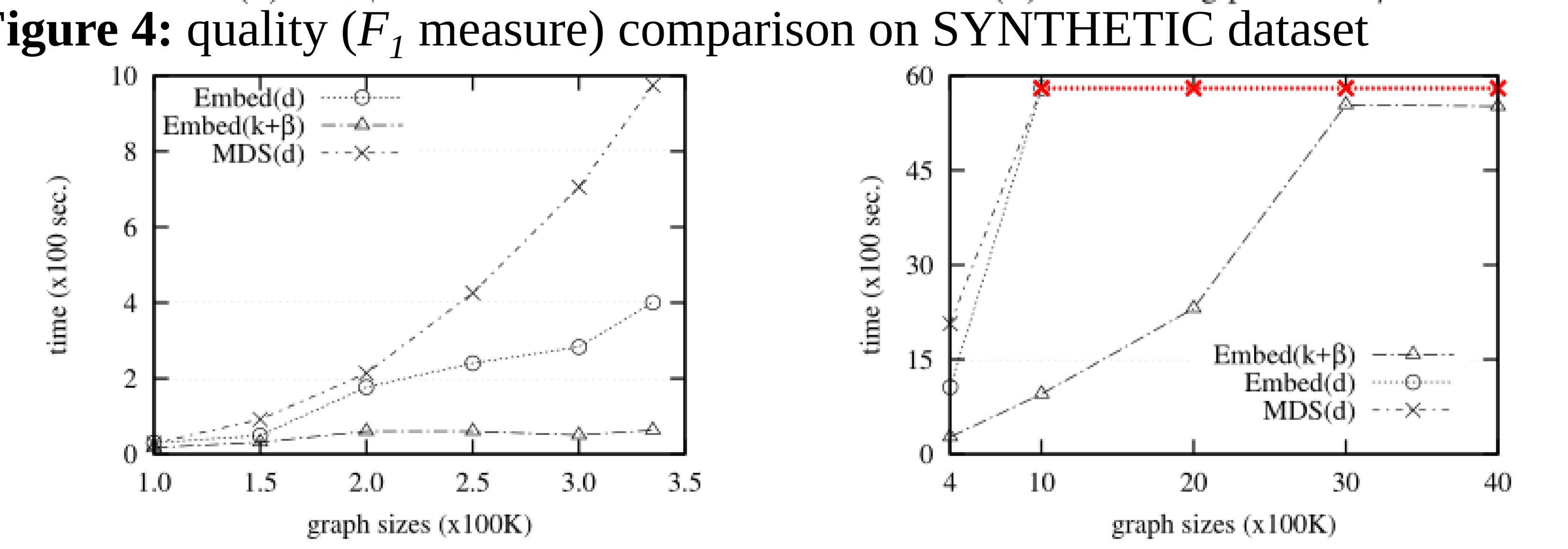
Figure 3: improvement on effectiveness of community detection (modularity)

The removal of detected anomalies helps improve the effectiveness of community detection.



(a) $k+\beta$ reduction (b) The mixing parameter μ

Figure 4: quality (F_1 measure) comparison on SYNTHETIC dataset



(a) AMAZON dataset (b) SYNTHETIC dataset

Figure 5: efficiency comparison *w.r.t* the graph sizes

Our embedding approach to anomaly detection is both effective and efficient. Moreover, the ($k+\beta$) reduction technique reduces space cost and improves efficiency in the same time, and slightly improves effectiveness.

References

[1] R. S. Burt. Structural holes and good ideas. *American Journal of Sociology*, 110(2): 349-399, 2004.
[2] M. McPherson, L. Simth-lovin, and J. Cook. Birds of a feather: Homophily in social networks. *Annual review of sociology*, Vol. 27: 415-444, 2001.