

1 Abstract

We present HARP, a novel method for learning low dimensional embeddings of a graph's nodes which preserves higher-order structural features. Our proposed method achieves this by compressing the input graph prior to embedding it, effectively avoiding troublesome embedding configurations (i.e. local minima) which can pose problems to non-convex optimization.

HARP works by finding a smaller graph which approximates the global structure of its input. This simplified graph is used to learn a set of initial representations, which serve as good initializations for learning representations in the original, detailed graph. We inductively extend this idea, by decomposing a graph in a series of levels, and then embed the hierarchy of graphs from the coarsest one to the original graph.

HARP is a general meta-strategy to improve all of the state-of-the-art neural algorithms for embedding graphs, including DeepWalk, LINE, and Node2vec. Indeed, we demonstrate that applying HARP's hierarchical paradigm yields improved implementations for all three of these methods, as evaluated on classification tasks on real-world graphs such as DBLP, BlogCatalog, and CiteSeer, where we achieve a performance gain over the original implementations by up to 14% Macro F1.

HARP는 그래프 노드의 저차원 임베딩을 학습하는 새로운 방법으로, 고차원 구조적 특징을 보존합니다. 저희가 제안하는 이 방법은 임베딩 전에 입력 그래프를 압축하여, 비볼록 최적화(non-convex optimization)에 문제를 일으킬 수 있는 골치 아픈 임베딩 구성(즉, 지역 최솟값)을 효과적으로 피함으로써 이를 달성합니다.

HARP는 입력 그래프의 전역적인 구조를 근사하는 더 작은 그래프를 찾는 방식으로 작동합니다. 이 단순화된 그래프는 초기 표현 집합을 학습하는 데 사용되며, 이는 원본의 상세한 그래프에서 표현을 학습하기 위한 좋은 초기값 역할을 합니다. 저희는 그래프를 일련의 수준으로 분해한 다음, 가장 coarse한(조밀도가 낮은) 그래프부터 원본 그래프까지 계층 구조의 그래프를 임베딩함으로써 이 아이디어를 귀납적으로 확장합니다.

HARP는 DeepWalk, LINE, Node2vec을 포함한 모든 최신 신경망 그래프 임베딩 알고리즘을 개선하기 위한 일반적인 메타 전략입니다. 실제로, DBLP, BlogCatalog, CiteSeer와 같은 실제 그래프에 대한 분류 작업 평가에서 HARP의 계층적 패러다임을 적용하면 이 세 가지 방법 모두에 대해 개선된 구현을 이끌어내며, 기존 구현 대비 최대 14%의 Macro F1 성능 향상을 달성함을 입증합니다.

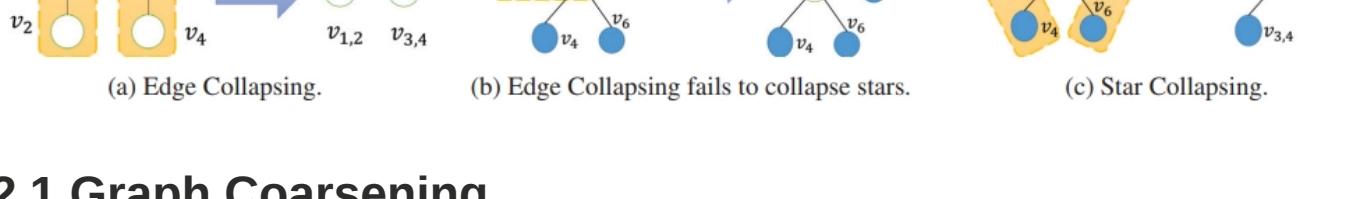
2 정리

dimensionality reduction이나 DeepWalk 등의 방법들은 몇 가지 공통된 약점을 가진다.

1. 노드의 지역적인 정보만 반영하고
2. SGD와 같은 방법을 사용하기 때문에 local minima에 빠지기 쉽다.

Coarsening - Embedding - Representation Prolongation and Refinement의 3가지 단계로 구성

1. Coarsening: 그래프의 global structure를 보존하는 적은 노드 수의 계층을 생성한다. (G_0, \dots, G_L)
2. Embedding: Coarsening된 그래프에 대해 embedding을 수행한다.
3. Representation Refinement: G_{i+1} 의 embedding을 G_i 의 초기값 Φ'_{G_i} 으로 사용한다.
4. 이를 반복하여 G_0 의 embedding을 얻는다.



2.1 Graph Coarsening

Edge Collapsing

중복되는 정점을 가지지 않는 간선들의 집합 E' 을 선택하고, $(u_i, v_i) \in E'$ 인 정점들을 합쳐 노드 w_i 를 만든다.

Star Collapsing

많은 실제 그래프가 scale-free network에 해당하며, hub를 가짐.

이러한 hub는 (b)와 같이 edge collapsing으로 표현하기 힘들기 때문에, 다른 방식 사용
유사한 이웃 노드를 가진 노드끼리 merge

2.2 Representation Prolongation and Refinement

Embedding Prolongation

supernodes를 포함하는 G_{i+1} 에 대한 embedding을 학습하고 난 뒤, 이를 초기값으로 하여 G_i 의 임베딩을 학습한다.

3 Conclusion

그래프에서 단순 edge와 star형 subgraph를 hierarchical하게 바꾸기 위한 방법을 제시

이러한 방법이 global structure에 대한 정보를 포함하는 embedding의 초기값을 제공하여, local minima에 빠지지 않도록 해줌.