Review of Network Abstraction Techniques

Fang Zhou, Sébastien Mahler, Hannu Toivonen

Department of Computer Science and Helsinki Institute for Information Technology HIIT, University of Helsinki

September 11, 2009
Workshop on Explorative Analytics of Information Networks
Review of Network Abstraction Techniques
1 Introduction

2 Structure of the field

3 Review of Techniques

4 Conclusions
Outline

1. Introduction
2. Structure of the field
3. Review of Techniques
4. Conclusions
Definition

What is *Network Abstraction*?

*Network Abstraction* is to extract a *simpler* but *more useful* graph from a large graph.

*Network Abstraction* helps the user to:

- observe the *essential* structure of a large network;
- identify *implicit* connections between distant nodes;
- discover *novel* knowledge.
Motivation


- **nodes**: 1 million
  - label: "gene", "protein", "article"...

- **edges**: 10 million
  - label: "codes for", "is homologous to"...
  - weight: the probability that the edge exists
Outline

1 Introduction

2 Structure of the field

3 Review of Techniques

4 Conclusions

Fang Zhou  
Review of Network Abstraction Techniques
Operations

Networks can be abstracted with three kinds of operations:

1. Pruning
2. Partitioning
3. Replacing
Operations I: Pruning

Pruning: to prune peripheral or irrelevant nodes and edges, this reduces the size of the network.
Operations I : Pruning
Example: original graph
Operations I : Pruning
Example: result
Operations II: Partitioning

Partitioning: to partition the network into smaller ones, and explore them individually.
Operations II: Partitioning

Example: original graph
Operations II: Partitioning

Example: partitioning
Operations II : Partitioning

Example: result
Operations III: Replacing

Replacing: to replace a part of the network by a more general structure.
Operations III: Replacing

Example: original graph
Operations III : Replacing

Example: result
Another axis of classification

A second way of classifying network abstraction methods:

- objective *versus* subjective methods
Objective *versus* Subjective Methods

Example: Original Network
Objective *versus* Subjective Methods

Example: Objective Method — keep best paths between all pairs of nodes
Objective *versus* Subjective methods

Example: Original Network
Objective *versus* Subjective methods

Example: Subjective Method — keep best paths between query nodes
Classification of abstraction methods

1. objective vs. subjective methods.

2. Three operations: pruning, partitioning and replacing.
Outline

1. Introduction
2. Structure of the field
3. Review of Techniques
4. Conclusions
Objective Methods

The objective methods abstract a graph using information only from the graph itself, such as its topology or edge weights.

Review of objective methods:

1. Pruning
2. Partitioning
3. Replacing
Objective Methods:
Pruning Edges or Nodes (1/5)

Relative Neighborhood Graph [1]:
it prunes edges between nodes $a$ and $b$ if they have a shared close node $c$;

[1] Toussaint, G.T.
Objective Methods:
Pruning Edges or Nodes (2/5)

To prune unimportant nodes: first rank the nodes.

- Ranking in an unweighted graph: use Node Centrality

  - Degree centrality:
    nodes with more edges are more central;

  - Betweenness centrality [7-9]:
    a node’s betweenness is the number of times the node appears on the paths between all other nodes;

Objective Methods: \textbf{Pruning Edges or Nodes (3/5)}

- Ranking in an unweighted graph: use Node Centrality (continued)
  - Closeness centrality \cite{11}:
    it is defined as the sum of graph-theoretic distances from a given node to all others in the network;
  - Feedback centrality:
    a node’s feedback centrality is defined recursively by the centrality of its adjacent vertices.

\cite{11} Gert, S.
Objective Methods: Pruning Edges or Nodes (4/5)

  Both of them discover the important web pages according to the web’s link structure.

---

[13] Lawrence, P. *et al.*
Objective Methods:
Pruning Edges or Nodes (5/5)

- Ranking in a probabilistic graph
  - Birnbaum's Component Importance [18]: The importance of an edge depends on the overall effect of the existence of the edge.

[18] Birnbaum, Z.W
Objective Methods: Partitioning a Graph (1/3)

Graph Partitioning:
The basic goal is to divide the nodes into subsets of roughly equal size and to minimize the sum of weights of edges crossing different subsets.

1 Spectral bisection [22,23]:
it bisects a graph after having sorted its nodes with respect to their Fiedler vector coordinates.

Objective Methods: Partitioning a Graph (2/3)

- Graph Partitioning: (continued)
  
  2 Geometric partitioning [24,25]: it first selects a coordinate axis, then finds a plane which is orthogonal to the selected axis, and bisects the nodes.

  3 Kernighan-Lin algorithm [29]: it takes a rough partitioning as input and iteratively swaps a subset of nodes in order to minimize the edge-cut.

  4 Multilevel method [26,27]: it first collapses some nodes and edges, then partitions and refines the partitioning while projecting back.

---

[29] Kernighan, B.W. et al.
Objective Methods:
Partitioning a Graph (3/3)

- Hierarchical Clustering (see e.g. [32]):
  an incremental approach: it optimizes a partition at one time.

- Edge Betweenness [33]:
  edge betweenness is the number of paths that run along that given edge, and it prunes some edges with high betweenness.

---

[32] Scott, J.
Objective Methods: 
Replacing Subgraphs (1/1)

- Clustering
  it replaces a dense cluster by a single node;

- Finding frequent Subgraphs: it can help us to replace frequent subgraphs in a large graph
  - approaches based on Inductive Logic Programming\cite{39};
  - methods based on Apriori algorithm\cite{43};
  - gSpan\cite{47} is based on depth-first search;
  - CloseGraph\cite{48} mines closed frequent graphs;
  - Spin\cite{49} mines maximal connected frequent subgraphs.

\cite{39} Dehaspe, L. \textit{et al.}
\cite{43} Agrawal, R. \textit{et al.}
\cite{47} Yan, X. \textit{et al.}
\cite{48} Yan, X. \textit{et al.}
\cite{49} Huan, J. \textit{et al.}
Subjective Methods
Subjective Methods

The subjective methods aim to maintain more information about those parts of a network that users are interested in.

Review of subjective methods:

1. Pruning
2. Partitioning
3. Replacing
Subjective Methods:
Pruning Edges or Nodes (1/2)

- Relevant Subgraph Extraction Problem

Several definitions of the quality of the extracted subnetwork:
- counting of edge-disjoint or vertex-disjoint paths from source to terminal [54];
- based on electricity analogies, aiming at maximizing electrical currents in a network [56];
- solving center-piece subgraph problem [57];
- computing network reliability [58].

[56] Faloutsos, C. et al.
[57] Tong, H. et al.
[58] Sevon, P. et al.
Subjective Methods:
Pruning Edges or Nodes (2/2)

- Interesting Nodes or Paths Detection [62]
  - using *rarity* to find interesting paths or nodes;
  - *rarity*: an event that occurs infrequently compared to other events, is potentially interesting and worth being reported.

- Personalized PageRank [64,65]
  It assigns importance according to the query or user preferences.

[62] Lin, S. *et al.*
Generic clustering methods that allow user input, such as constrained clustering [66] or supervised clustering [67], could be applied to graphs.

[67] Eick, C.F. et al.
Subjective Methods: Replacing Subgraphs (1/2)

Finding subgraphs methods: it can help us to replace the user’s input by general nodes.

- **Exact search**
  - subgraph isomorphism problem: NP-complete
  - how to number the isomorphisms?
  - how to index a database of graphs: GraphGrep [70], GIndex [71].

[70] Shasha, D. *et al.*
[71] Yan, X. *et al.*
Subjective Methods:
Replacing Subgraphs (2/2)

- Similarity search
  - two kinds of similarity search:
    - K-NN (K-Nearest-Neighbors);
    - range query (subgraphs within a specific dissimilarity range).
  - similarity measures:
    ① edit distance [72];
    ② based on maximum common subgraph or minimum common subgraph [74,75];
    ③ maximum percentage of edges in common [76].

[72] He, H. et al.
[76] Yan, X. et al.
Outline

1. Introduction
2. Structure of the field
3. Review of Techniques
4. Conclusions
Conclusions

We review network abstraction approaches along two axes:

1. objective vs. subjective methods.

2. three types of operations: pruning, partitioning and replacing;
Conclusions: Open questions in the field

1. How to make the methods efficient when computing a large network?

2. How to involve the user in the partitioning process?

3. How to make replacing frequent subgraphs more feasible?

4. How to use subjective relevance to guide the replacement?

5. How to take advantage of these approaches to make an integrated approach?

6. How to define a global measure to value the quality of an abstracted network?
<table>
<thead>
<tr>
<th>Introduction</th>
<th>Structure of the field</th>
<th>Review of Techniques</th>
<th>Conclusions</th>
</tr>
</thead>
</table>

Fang Zhou

Review of Network Abstraction Techniques

thanks for your time!