Lecture VI - NP Problems: Cliques, Sets and Paths

[1] Department of Mathematics and Systems Analysis,
Systems Analysis Laboratory, Aalto University, Finland

March 1, 2024

Aalto University

Previously on..

- P vs NP
- TSP

PREVIOUSLY ON...

Cliques

Definition

Clique

A **clique** is a subset of nodes in a graph where every two distinct node is adjacent (e.g. there is an edge connecting every pair of nodes in the subset).

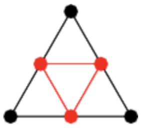


Figure: Example of a clique of consisting of three nodes.

History

It dates back to 1940s with **Turan's theorem** regarding clique in dense graphs. However, it gain popularity from the work of **Luce** & **Perry** in 1949 in social networks application.

Most effort in developing approximations and relaxations are present in the field of **communication networks** (starting from Prihar in 1959 and in following decades) and in **Bioinformatics** (by Ben-Dor, Shamir & Yakhini, 1999).

Applications

Mostly in Communication:

- Design of efficient circuits;
- Automatic test patter generation;
- Hierarchical partition.

In Bioinformatics:

- Gene expression;
- Ecological niches;
- Metagenomics and evolutionary tree;

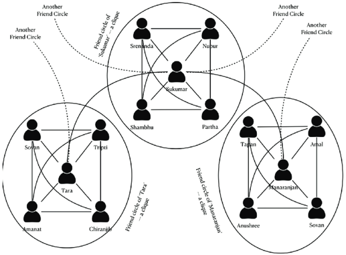


Figure: Clique application in telecommunication

Modelling

Aalto University

In its decision version, as well as several other versions (such as maximum clique, maximum clique with weights), is a **NP-hard problem**,.

For maximum clique problem (the largest possible clique in a graph), there is an ILP formulation.

Modelling - Variable and Objective

For each node $v \in V$, the decision variable x_v is assigned.

$$x_v = \begin{cases} 1, & \text{if node } v \text{ is selected as part the maximum clique} \\ 0, & \text{otherwise} \end{cases}$$

For objective function, the goal is to minimize the number of nodes selected as part of the maximum clique:

$$\max \sum_{v \in V}^{n} x_v$$

Remark: It can be further changed into **maximum weighted clique** by multiply the weight w_v to each selected node.

Modelling - Constraint

For constraint, a new concept is required to be introduced.

Let $\overline{E}$ be the **complementary edges** from G, such that:

- $(u, v) \notin \overline{E}$ if (u, v) in G;
- $(u, v) \in \overline{E}$, if (u, v) is not in G.

Hence, the constraint for maximum clique requires that for each edge $(u, v) \in \overline{E}$, at **least one** of the nodes **has** to be selected:

$$x_u + x_v \leq 1 \qquad \forall (u, v) \in \overline{E}$$

Modelling - Full Model

Aalto University

Maximize $\sum_{v \in V} x_v$

Subject to:

$$x_u + x_v \leq 1 \qquad \forall (u, v) \in \overline{E}$$

$$x_v \in \{0, 1\} \qquad \forall v \in V$$

Independent Set

Definition

Independent Set (IS)

An **independent set** is a subset of nodes in a graph where **there is no any adjacency** between all pair of nodes.

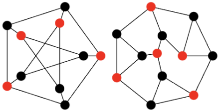


Figure: Example of independent sets

It is frequently called the anti-clique problem

History and Applications

Most of effort put in this problem is related to studies on **clique problems** and **vertex cover**.

However, some special version were studies by Perrin in the late 1890s and by Padovan in the 1990s.

Its applications are also related to clique problems, with emphasis stable genetic regions.

Modelling

Known as the **anti-clique**, the same models for cliques can be used.

$\longrightarrow$ requires used of complementary edges.

Definition

A set of nodes in a graph $G = (V, E)$ creates a clique if the same set form as an independent set in the complimentary graph $\overline{G} = (V, \overline{E})$.

Definition

A set is independent if and only if it is complement is a vertex cover.

Modelling - Variable and Objective

Alternatively, it also has an ILP formulation.

For each node $v \in V$, the decision variable x_v is assigned.

$$x_v = \left\{ \begin{array}{ll} 1, & \text{if node } v \text{ is selected as part an independent set} \\ 0, & \text{otherwise} \end{array} \right\}$$

For objective function, the goal is to maximize the number of nodes selected as part of an independent set:

$$\max \sum_{v \in V}^{n} x_v$$

Modelling - Constraint

If a node is part of an independent set, it cannot be adjacent to any other node in the independent set. Hence,

$$x_u + x_v \leq 1 \qquad \forall (u, v) \in \overline{E}$$

Modelling - Full Model

Aalto University

Maximize $\sum_{v \in V} x_v$

Subject to:

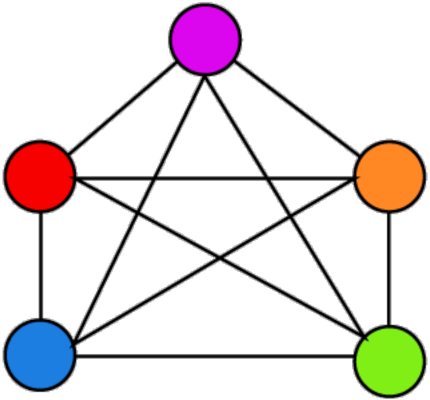
$$x_u + x_v \leq 1 \qquad \forall (u,v) \in E$$

$$x_v \in \{0,1\} \qquad \forall v \in V$$

Dominant Set

Definition

Dominant Set (DS)
A **dominant set** is subset of nodes from a graph $G = (V, E)$ such any node $v \in V$ is either part of the subset or adjacent to it.



History and Application

The **domination problem** (in various forms and extension) has been present in the literature since 1950s, with increased efforts in the 1970s due to its application in **networking** and **electrical grids**.

In **wireless network**, approximation algorithms have been used to find near-optimal routes withing ad-hoc networks.

Further applications are related to **document summarization** and **safety** in electrical grids.

Modelling - Variable and Objective Function

For each node $v \in V$, the decision variable x_v is assigned.

$$x_v = \left\{ \begin{array}{ll} 1, & \text{if node } v \text{ is selected as part an independent set} \\ 0, & \text{otherwise} \end{array} \right\}$$

For objective function, the goal is to maximize the number of nodes selected as part of a dominant set:

$$\min \sum_{v \in V}^{n} x_v$$

Modelling - Constraint

Aalto University

In a graph $G = (V, E)$, every node has two roles to play:

- It is part of the dominant set;
- It is adjacent to a node in the dominant set.

Those two roles are **mutually exclusive**: only of one can be satisfied for each node v in V. This can be modelled by the boolean operator OR.

Modelling - Constraint

Breaking down the two roles:

- It is part of the dominant;

$$x_v = 1 \qquad \forall v \in V$$

- It is adjacent to a node in the dominant set.

$$\sum_{(u,v)\in E} x_u = 1 \qquad \forall v \in V$$

Applying the boolean operator results in the following constraint:

$$x_v + \sum_{(u,v)\in E} x_u \geq 1 \qquad \forall v \in V$$

Modelling - Full Model

Minimize $\sum_{v \in V} x_v$

Subject to:

$$x_v + \sum_{(u,v) \in E} x_u \geq 1 \qquad \forall v \in V$$

$$x_v \in \{0, 1\} \qquad \forall v \in V$$

Hamiltonian Path

Definition

Hamiltonian Path/Cycle
A Hamiltonian path/cycle (or traceable path/circuit) is a path in a graph where **every node is visited only once**.

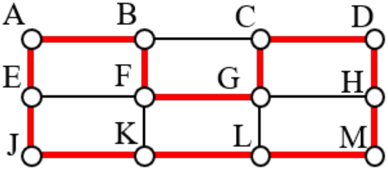


Figure: Example of a hamiltonian path

History

Its name is associated to **William Rowan Hamilton**, inventor of icosian game (finding a cycle in the edge graph of a dodecahedron - also known as Hamilton's puzzle).

However, this problem have been studied earlier by Thomas Kirkman and similar problems have been address by **Indian** and **Islamic mathematician** since the 9th century.

One of the most **famous NP-hard problem**, it has been widely research since the 1972, when concrete solutions were proposed by Bondy-Chvatal by on previous work by Dirac and Ore in 1952.

Application

Many areas with interest in TSP also research Hamiltonian paths, with emphasis in:

- Routing problems, such as vehicle routing problem;
- Electronic circuit design;
- Computer graphics;
- Genome mapping.

Modelling - Variable and Objective

In both Hamiltonian path or cycle formulation, the decision variable x_{uv} is assigned to each edge (u, v).

$$x_{u,v} = \left\{ \begin{array}{ll} 1, & \text{if edge } (u,v) \text{ is part of a Hamiltonian path or cycle} \\ 0, & \text{otherwise} \end{array} \right\}$$

Our objective function is to minimise the cardinality of such path or cycle:

$$\min \sum_{(u,v) \in E} x_{u,v}$$

Modelling - Constraint

For **both cycle and path**, it is required that each node is visited by two edges. Then, the following constraint:

$$\sum_{v \in V} x_{u,v} = 2 \qquad \forall u \in V$$

Modelling - Constraint - HC

For Hamiltonian cycles, for any subset of nodes that is not V or $\emptyset$, there should be at least two edges being used, resulting in the constraint:

$$\sum_{(u,v):u \in S, v \notin S} x_{u,v} \geq 2 \qquad \forall S \subset V, S \neq 0$$

Modelling - Constraint - HP

For Hamiltonian path, both **source** $s \in V$ and **sink** $t \in V$ should be selected (assuming single source and sink). For any intermediary node, the same amount of incoming edges should be equal to outgoing edges.

This results in the following constraints:

$$\sum_{(uv)\in\delta^+(u)} x_{u,v} - \sum_{(v,u)\in\delta^-(u)} x_{u,v} = b_u \qquad \forall u \in V$$

where $\delta^+(u)$ and $\delta^-(u)$ represents the **out-** and **in-edges** from u. b_v assumes

different values if v is a **source** $b_v = 1$, a **sink** $b_v = -1$ or an **intermediary** node $b_v = 0$

Modelling - Constraint - HP

Two extra constraint are required:

- Only one out-edge is used:

$$\sum_{(u,v)\in\delta^+(u)} x_{u,v} \leq 1 \qquad \forall u \in V$$

- To impose that no subtour is present.

$\longrightarrow$ Similar subtour elimination from TSP.

Modelling - Full Model - HC

Aalto University

Minimize $\sum_{(u,v)\in E} x_{u,v}$

Subject to:

$$\sum_{v\in V} x_{u,v} = 2 \qquad \forall u \in V$$

$$\sum_{(u,v):u\in S,v\notin S} x_{uv} \geq 2 \qquad \forall S \subset V, S \neq 0$$

$$x_{u,v} \in \{0,1\} \qquad \forall v \in V$$

Modelling - Full Model - HP

Minimize $\sum_{(u,v)\in E} x_{u,v}$

Subject to:

$$\sum_{v\in V} x_{u,v} = 2 \qquad \forall u \in V$$

$$\sum_{(u,v)\in\delta^+(u)} x_{u,v} - \sum_{(v,u)\in\delta^-(u)} x_{u,v} = b_u \qquad \forall u \in V$$

$$\sum_{(u,v)\in\delta^+(u)} x_{uv} \leq 1 \qquad \forall u \in V$$

$$x_{u,v} \in \{0,1\} \qquad \forall v \in V$$

Modelling - Full Model - HP

- Require at least one extra constraint for subtour elimination; Either DFJ or MTZ (as examples);
- ${}^{*}b_u \in \{-1, 0, 1\}$; For all nodes $v \in V$ depending if they are a source, a sink or an intermediary nodes.

Conclusion

Challenges

All new problems have two elements in common:

- Hard/impossible to solve in polynomial time;
- Easy to verify a solution in polynomial time.

Also, all these problems can be easily "transformed" into each other.

Class NP

All these problems are **NP problems**.
But how do you **prove** a novel problem is NP?

Is there any **correlation** between?

Any form of **transformation**?


thank you