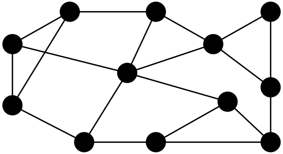
Intractable Problems

- Clique
- Independent Set
- Vertex Cover
- Set Cover
- Set Packing
- Satisfiability Problem
- Hamiltonian Cycle and Path
- Traveling Salesman Problem
- Graph Coloring
- Circuit Satisfiability
- Knapsack
- Subset Sum
- Prime and Factor
- Partition

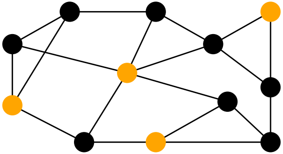
Imdadullah Khan

Independent Set in Graph

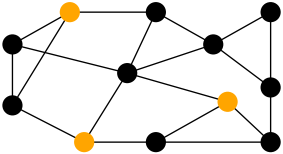
An independent set in G is subset of vertices no two of which are adjacent



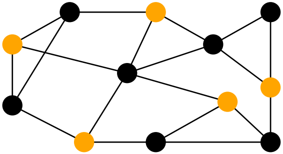
A graph on 12 vertices



An independent set of size 4



An independent set of size 3



An independent set of size 5 (max)

The IND-SET(G, k) problem: Is there an independent set of size k in G?

Independent Set Applications

Sites Selection Problem

- Suppose n potential sites are identified for opening up restaurants
- Some pairs of places shouldn't have the franchises at both of them
  - too close to each other, competitions, or operational constraints
- Make a graph G with vertices as sites and edges as pairwise conflicts
- Selecting k sites becomes finding a k-independent set in G

Independent Set Applications

The SNP (Single Nucleotide Polymorphism) Assembly Problem

- In computational biology (biochemistry) given a set of sequences we want to resolve inter-sequential conflicts by excluding some sequences
- Conflict between two sequences is due to their biochemical properties
- The goal is to select a large number of conflict free sequences
- Make a graph with vertices representing sequences and edges representing conflicts
- Find a large independent set in this graph

Independent Set Applications

Diversifying Investment Portfolio

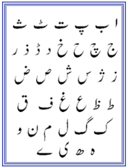
- Different stocks in a market
- $P_i(t)$ is price for stock i at time t
- $R_i(t) = \log \dfrac{P_i(t)}{P_i(t-1)}$, return or trading volume of stock i at time t
- Make each stock a node and two stocks have edges if correlation of their returns is $\geq \theta$ for threshold $-1 \leq \theta \leq 1$
- θ is set depending on potential risk (degree of diversification)
- Two adjacent vertices in $G_{\theta=.9}$ represent high risk investment pair

Set $\theta < -0.5$: an independent set in G_θ represents a portfolio with "small" risk (diverse set of investments)

Independent Set Applications

Shannon Capacity of a graph

- Sending a message from an alphabet through a noisy channel

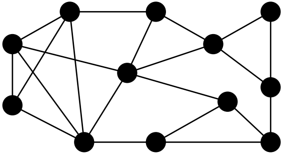


- Because of noise some characters can be confused
- How many 1 length strings can be sent without confusion?
- Make each letter a node and make edges iff the corresponding letters can be confused (depends on the SNR of channel)
- Max number of messages is the size of max independent set

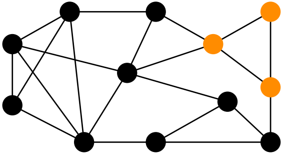
- How many k-length strings can be sent on this channel?
- Size of max independent set in G^k (strong product of graphs)

Cliques in Graphs

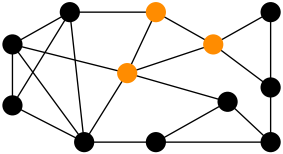
A **clique** in G is a subset of vertices every two of which are adjacent



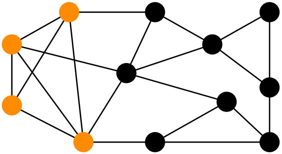
A graph on 12 vertices



A clique of size 3



A clique of size 3



A clique of size 4 (max)

The $\textsc{clique}(G, k)$ problem: Is there a clique of size k in G?

Clique Applications

Cliques in Market Graphs

- Different stocks in a market
- $P_i(t)$ is price for stock i at time t
- $R_i(t) = \log \frac{P_i(t)}{P_i(t-1)}$, return or trading volume of stock i at time t
- Each stock is a node and two stocks have edges if correlation of their returns is $\geq \theta$ for threshold $-1 \leq \theta \leq 1$
- θ is set depending on potential risk (degree of diversification)
- Two adjacent vertices in $G_{\theta=.9}$ represent high risk investment pair

Set $\theta > 0.5$: a clique in G_θ represents a portfolio with "large" risk

Can also be of interest to a regulatory body to determine collusion

Clique Applications

Organized Tax Fraud Detection by IRS

- Clustering similar objects is widely used in many applications
- Ideal clusters are cliques in a graph (community, highest internal degrees, lowest internal distances, largest internal densities etc.)
- Groups of phony tax returns are submitted to get undeserved returns
- IRS constructed graph, where each returned form is a vertex
- Edges between two vertices means '*similarity* between the two forms is above a certain threshold
- A large clique in this graph points to a potential fraud

Location Covering Using Clique Partition

Protein Docking Problem