Harvard CS 121 and CSCI E-207

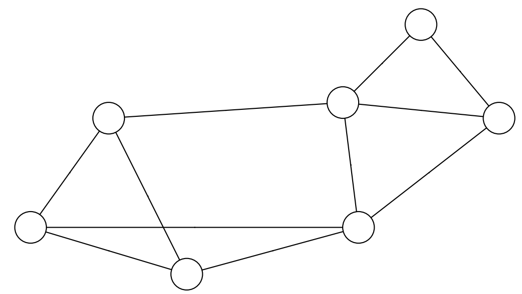
Lecture 23: More NP-completeness

Salil Vadhan

November 27, 2012

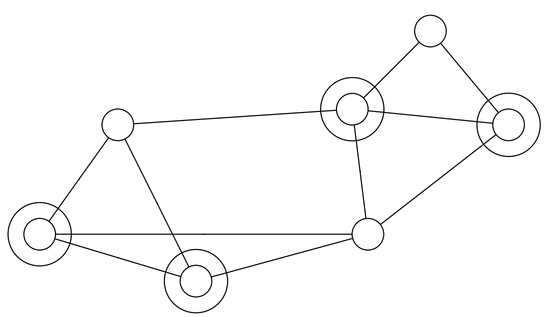
VERTEX COVER (VC)

- Instance:
  - a graph, e.g.



  - a number k (e.g. 4)

- Question: Is there a set of k vertices that "cover" the graph, i.e., include at least one endpoint of every edge?



VC is NP-complete

- VC is in NP:

- 3-SAT $\leq_{\mathrm{P}}$ VC:

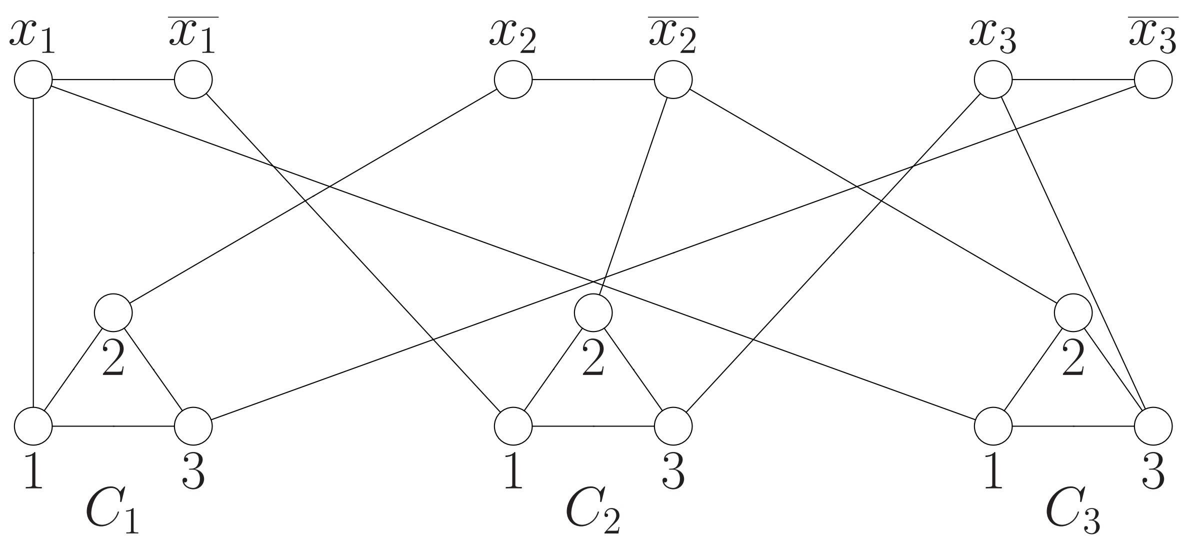
  - Let F be a 3-CNF formula with clauses $C_1 \ldots, C_m$, variables $x_1, \ldots, x_n$.

  - We construct a graph G_F and a number N_F such that:

    G_F **has a size** N_F **vertex cover iff** F **is satisfiable**

Construction of G_F and N_F from F

E.g. $F = (x_1 \vee x_2 \vee \neg x_3) \wedge (\neg x_1 \vee \neg x_2 \vee x_3) \wedge (x_1 \vee \neg x_2 \vee x_3)$



- G_F = one dumbbell for each variable, one triangle for each clause, and corner j of triangle i is connected to the vertex representing the jth literal in C_i.

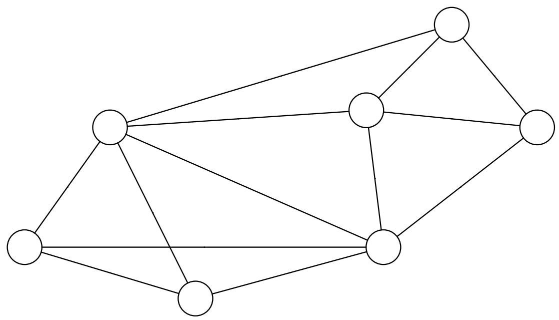
- $N_F = 2m + n = 2$ (# clauses) $+$ (# variables).
$\Rightarrow$ 1 vertex from each dumbbell and 2 from each triangle.

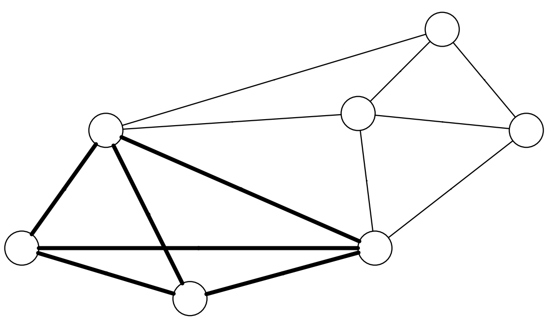
Correctness of the Reduction

- If F is satisfiable, then there is an N_F cover:

- If there is an N_F cover, then F is satisfiable:

CLIQUE

- Instance:
  - a graph, e.g. 
  - a number k (e.g. 4)
- Question: Is there a clique of size k, i.e., a set of k vertices such that there is an edge between each pair?



- Easy to see that CLIQUE $\in$ NP.

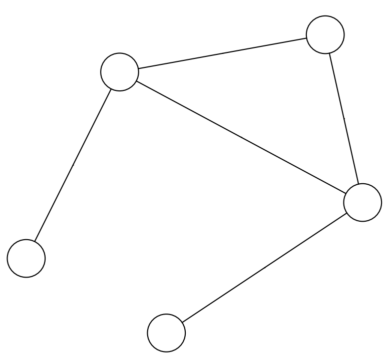
VC $\leq_P$ CLIQUE

If G is any graph, let G^c be the graph with the same vertices such that:

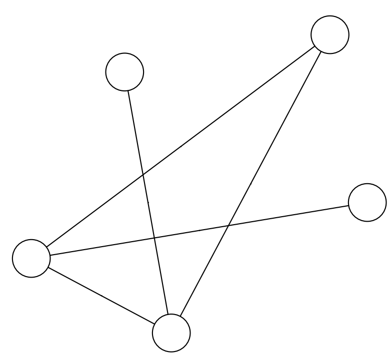
there is an edge between x and y in G^c
iff
there is no edge between x and y in G

e.g.

$G =$



$G^c =$



VC $\leq_P$ CLIQUE, continued

Let (G, k) be an instance of VC.

Claim: G has a k-cover iff G^c has a $|G| - k$ clique, where $|G|$ is the number of vertices in G.

(So the mapping $(G, k) \mapsto (G^c, |G| - k)$ is a reduction of VC to CLIQUE.)

Proof:

INTEGER LINEAR PROGRAMMING

An integer linear program is

- A set of variables $x_1, \ldots, x_n$ which must take integer values.
- A set of linear inequalities:

$$a_{i1}x_1 + a_{i2}x_2 + \ldots + a_{in}x_n \le c_i \qquad [i = 1, \ldots, m]$$

e.g. $x_1 - 2x_2 + x_4 \le 7$

$x_1 \ge 0 \qquad [-x_1 \le 0]$

$x_4 + x_1 \le 3$

ILP $=$ the set of integer linear programs for which there are values for the variables that simultaneously satisfy all the inequalities.

ILP is NP-complete

INTEGER LINEAR PROGRAMMING $\in$ NP. (Not obvious! Need a little math to prove it. Proof omitted.)

INTEGER LINEAR PROGRAMMING is NP-hard: by reduction from 3-SAT $($3-SAT $\leq_{\text{P}}$ ILP$)$. Given 3-CNF Formula F, construct following ILP P as follows:

Note: LINEAR PROGRAMMING where the variables can take *real* values is known to be in P.

More NP-complete/NP-hard Problems

- HAMILTONIAN CIRCUIT (and hence TSP) (see Sipser for related problems)
- SCHEDULING
- CIRCUIT MINIMIZATION
- SHORT PROOF
- NASH EQUILIBRIUM WITH MAXIMUM PAYOFF
- PROTEIN FOLDING
- ⋮
- See Garey & Johnson for hundreds more.