set cover

```
  * instance  universe U = {1, 2, ..., n}
              set S of m subsets of U
    query     are there k sets in S that cover U ?

  * also called min set cover problem       (why?)

  * NP-complete                              (why?)

         set     universe
             0 1 2 3 4 5 6 7 8 9
         0   - - * - - - - * * -
         1   - * - - - - - * - *
         2   - - - - * - - * * -
         3   - - - - - - - - * -
         4   - - - - - * - * * -
         5   - - - - * - * * - *
         6   * * - * - - - - - -
         7   - - - - - - * * - -

   is this a cover?
         0   - - * - - - - * * -
         4   - - - - - * - * * -
         5   - - - - * - * * - *
         6   * * - * - - - - - -
   is this a min cover?
```

```
************************
 greedy set cover alg'm
************************

C <- { }
loop
    Z <- all nodes not covered by C
    if Z empty:
        exit loop
    for each c in S\C:
        c_z <- size of intersection c,Z
    add any c with max c_z to C
return C

************************

previous example:
          5   - - - - * - * * - *
          6   * * - * - - - - - -
          0   - - * - - - - * * -
          4   - - - - - * - * * -
```

****

how bad can greedy set cover be ?

****

Theorem. n-node GSC has size $\le opt * (\ln n)$

Proof. n_t: number uncovered nodes after t it'ns

* $n_0 = n$

* $n_1 \le n_0 - n_0/opt = n_0 (1 - 1/opt)$ why?

* opt-cover: some set that hits n/opt nodes

* $n_{t+1} \le n_t - n_t/opt = n_t (1 - 1/opt)$ why?

* opt-cover: some set that hits n_t/opt nodes

  (if greedy cover-so-far has t sets from opt-cover,

  then some remaining set hits $n_t/(opt - t)$ nodes)

* $n_t \le n_0 (1 - 1/opt)^t < n_0 (e^{-1/opt})^t =$

  $ne^{-t/opt}$

* when $t = opt \ln n$, $n_t < n e^{-\ln n} = 1$ Q.E.D.

worst-case lower bound

```
                                      n  2
    . *                       min cover  2
    * .                    greedy cover  2
```

```
  . . * *
  . * . .
  * . . .                             n  4
  . * . *                     min cover  2
  * . * .                  greedy cover  3
```

```
. . . . * * * *
. . * * . . . .
. * . . . . . .
* . . . . . . .                       n  8
. * . * . * . *               min cover  2
* . * . * . * .            greedy cover  4
```

```
. . . . . . . . * * * * * * * *
. . . . * * * * . . . . . . . .
. . * * . . . . . . . . . . . .
. * . . . . . . . . . . . . . .
* . . . . . . . . . . . . . . .       n 16
. * . * . * . * . * . * . * . *   min cover  2
* . * . * . * . * . * . * . * .   greedy cover  5
```

Claim: for all $t \geq 1$, for $n = 2^t$,

there is a set cover instance with

min cover size 2

greedy cover size $t + 1$.

Corollary: for all $t \geq 1$, for $n = 2^t$,

there is a set cover instance with

(greedy cover size) / (min cover size) in

$\Omega(\lg n)$.

Corollary: for all n,

worst case greedy set cover / optimal set cover

is in $\Theta(\lg n)$