

Linguagens Formais e Autômatos

Aula 28 - Problemas P e NP

Prof. Dr. Daniel Lucrédio
Departamento de Computação / UFSCar
Última revisão: ago/2015

Referências bibliográficas

- **Introdução à teoria dos autômatos, linguagens e computação / John E. Hopcroft, Rajeev Motwani, Jeffrey D. Ullman** ; tradução da 2.ed. original de Vandenberg D. de Souza. - Rio de Janeiro : Elsevier, 2002 (Tradução de: Introduction to automata theory, languages, and computation - ISBN 85-352-1072-5)
 - Capítulo 10 - Seção 10.1
- **Introdução à teoria da computação / Michael Sipser** ; tradução técnica Ruy José Guerra Barretto de Queiroz ; revisão técnica Newton José Vieira. -- São Paulo : Thomson Learning, 2007 (Título original : Introduction to the theory of computation. "Tradução da segunda edição norte-americana" - ISBN 978-85-221-0499-4)
 - Capítulo 7 - Seções 7.2 e 7.3

Introdução

- Nessa aula, iremos partir da análise sobre:
 - o que pode ser resolvido por um computador (e o que não pode)
- E passaremos a analisar:
 - o que pode ser resolvido por um computador em um tempo realístico (e o que não pode)
- Ou seja, o fato de um problema ser decidível não é suficiente
 - Pois ele pode levar um longo tempo
 - Ou pode exigir muito espaço em memória

O que é “muito tempo”?

- Qual a linha divisória entre problemas tratáveis e problemas intratáveis?
- Na teoria da complexidade/intratabilidade
 - Tempo polinomial (tratáveis)
 - Tempo exponencial (intratáveis)
- Na realidade, existem problemas com tempo intermediário
 - Ex: $n^{\log_2 n}$
 - Esses também são intratáveis
- Na prática, a maioria dos problemas se enquadra em polinomial ou exponencial

O que é “muito tempo”?

- Para todas análises que se seguem:
 - Diferenças polinomiais em tempo de execução são consideradas pequenas
 - Diferenças exponenciais em tempo de execução são consideradas grandes
- Razão: a diferença entre a taxa de crescimento de polinômios e exponenciais
 - Ex: n^3 e 2^n , para $n = 1000$
 - $n^3 = 1000000000$ é grande, mas não muito, em termos computacionais
 - 2^n = mais do que o número de átomos no universo

Algoritmos exponenciais

- Normalmente não tem utilidade prática
 - Conseguem chegar a uma decisão apenas para instâncias MUITO pequenas
- Tipicamente envolvem busca exaustiva
 - Busca pela força bruta
 - Ex: fatorar um número em seus primos, buscando-se todos os potenciais divisores
 - Às vezes, pode-se evitar a busca por força bruta
 - Entendendo-se melhor o problema, revelando um algoritmo de tempo polinomial

Diferenças polinomiais

- Todos os modelos computacionais determinísticos são polinomialmente equivalentes
 - Máquina de Turing determinística com uma fita
 - Máquina de Turing determinística com múltiplas fitas
 - Computador real
 - Outros dispositivos determinísticos semelhantes
- Peraí:
 - Programadores trabalham duro para tentar fazer programas rodarem duas vezes mais rápido
 - E você está dizendo que rodar em n ou n^{1000} é equivalente?

Diferenças polinomiais

- Na prática, diferenças polinomiais nem sempre são ignoráveis
- Mas essas diferenças tendem a serem resolvidas com o tempo
- Ao passo que para diferenças SIGNIFICATIVAS não há perspectiva iminente de solução
 - Como por exemplo: polinomial X exponencial
 - É justamente pela diferença na magnitude citada anteriormente

A classe P

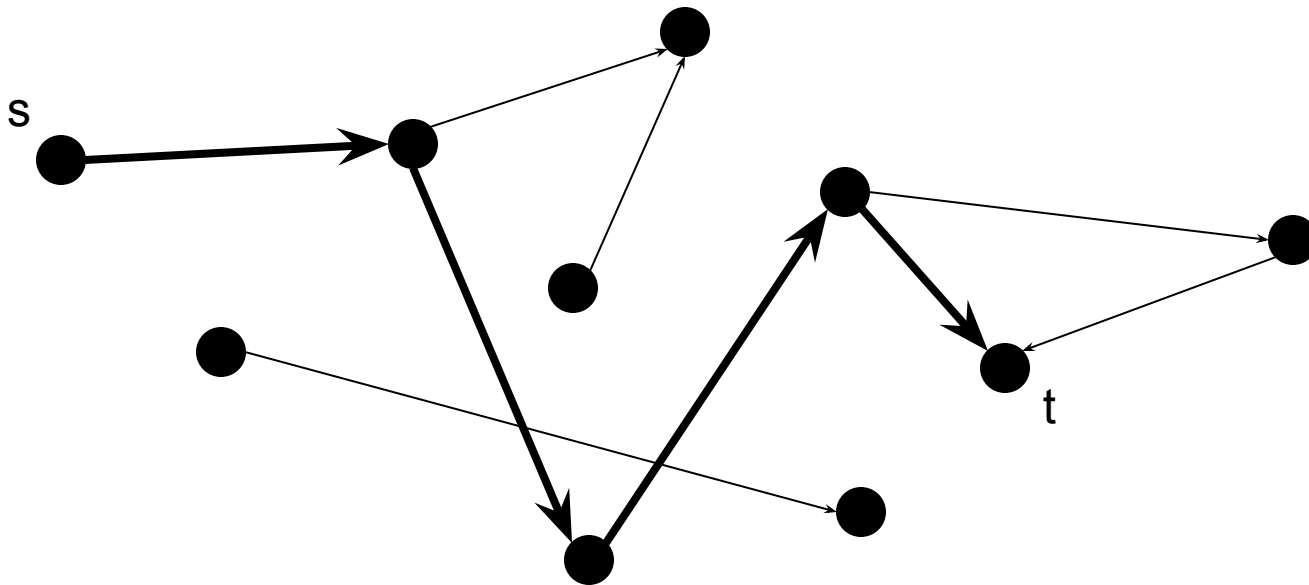
- Portanto, definimos uma importante classe de linguagens:

P é a classe de linguagens que são decidíveis em tempo polinomial sobre uma máquina de Turing determinística de uma fita

- P é invariante para todos os modelos de computação polinomialmente equivalentes à DTM de uma fita
- P corresponde aproximadamente à classe de problemas que são realisticamente solúveis em um computador

CAM

$CAM = \{(G,s,t) \mid G \text{ é um grafo direcionado, } s \text{ e } t \text{ são nós de } G, \text{ e } G \text{ tem um caminho direcionado de } s \text{ para } t\}$



CAM

- Existe um algoritmo exponencial que decide CAM
 - Usando busca pela força bruta, analisando todos os caminhos que passam por no máximo todos os nós de G , e encontrando um que envolve s e t
 - Guarde esse algoritmo na memória (usaremos depois)
- Mas existe um algoritmo de tempo polinomial que decide CAM
 - Ou: Existe uma MT determinística que sempre para e decide CAM
 - Portanto CAM pertence a P

CAM

Um algoritmo polinomial para CAM:

1. Ponha uma marca sobre o nó s
2. Repita o seguinte até que nenhum nó adicional seja marcado:
 3. Faça uma varredura em todas as arestas de G . Se uma aresta (a,b) for encontrada indo de um nó marcado a para um nó não-marcado b , marque o nó b
4. Se t estiver marcado, aceite. Caso contrário, rejeite.

PRIM-ES

- Dois números são primos entre si se 1 é o maior inteiro que divide ambos.
 - Ex: 10 e 21 são primos entre si, embora nenhum deles seja um número primo
 - Ex: 10 e 22 não são primos entre si, pois ambos são divisíveis por 2

$\text{PRIM-ES} = \{(x,y) | x \text{ e } y \text{ são primos entre si}\}$

PRIM-ES

- Existe um algoritmo exponencial
 - Busca por força bruta todos os possíveis divisores de ambos os números e aceita se nenhum é maior que 1
- Existe um algoritmo melhor, chamado algoritmo euclidiano:
 1. Repita até que $y=0$
 - a. Atribua $x \leftarrow x \bmod y$
 - b. Intercambie x e y
 2. Dê como saída x
 3. Se x for 1, aceite. Caso contrário, rejeite

A classe NP

- Evitar a força bruta é o objetivo
 - Muitas vezes permitem sair do exponencial e chegar no polinomial
 - Mas existem muitos (MUITOS) problemas interessantes, práticos e úteis, para os quais ainda não se conseguiu evitar a força bruta
 - Não se sabe se existem algoritmos de tempo polinomial que os resolvem

A classe NP

- Por que ninguém ainda encontrou algoritmos de tempo polinomial para esses problemas?
 - Não sabemos a resposta para essa questão!!
 - Talvez esses problemas tenham algoritmos de tempo polinomial, mas ninguém ainda descobriu! Talvez exista algum princípio desconhecido necessário.
 - Ou talvez, alguns desses problemas simplesmente **NÃO PODEM** ser resolvidos em tempo polinomial. Eles podem ser intrinsecamente difíceis.

A classe NP

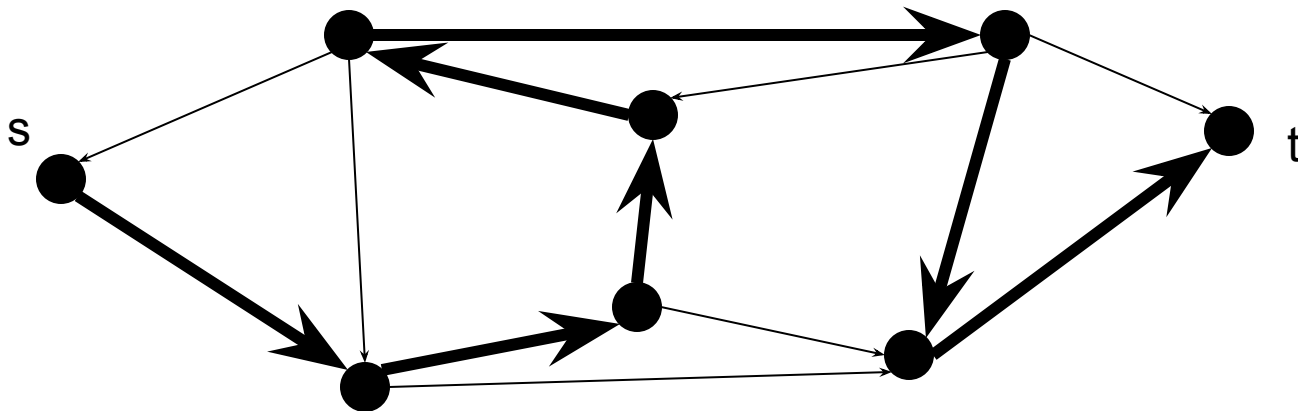
- Ou seja, aparentemente, existe uma classe, chamada NP, que é diferente de P (na verdade, NP é superconjunto de P)
 - Onde residem muitos problemas que não possuem solução polinomial
- Importante repetir: esses problemas têm uma solução, eles são decidíveis, existe um decisor!!!
 - Mas simplesmente, a solução (algoritmo) leva tempo polinomial
 - Isto significa, na prática, que uma entrada de tamanho razoável, como 1.000.000.000, por exemplo, pode levar um tempo equivalente à idade conhecida do universo

A classe NP

- Mas sabemos pelo menos UMA coisa sobre a classe NP
 - Os problemas em NP possuem uma característica em comum
 - Todos são **verificáveis** em tempo polinomial
 - Ou seja, uma vez sabendo a resposta, é fácil (a.k.a. Existe um algoritmo de tempo polinomial) verificar que a resposta é válida
- Veremos alguns exemplos para ilustrar esse fenômeno

CAMHAM

- Um caminho hamiltoniano em um grafo direcionado G é um caminho direcionado que passa por cada nó exatamente uma vez
- $\text{CAMHAM} = \{(G,s,t) \mid G \text{ é um grafo direcionado com um caminho hamiltoniano de } s \text{ para } t\}$

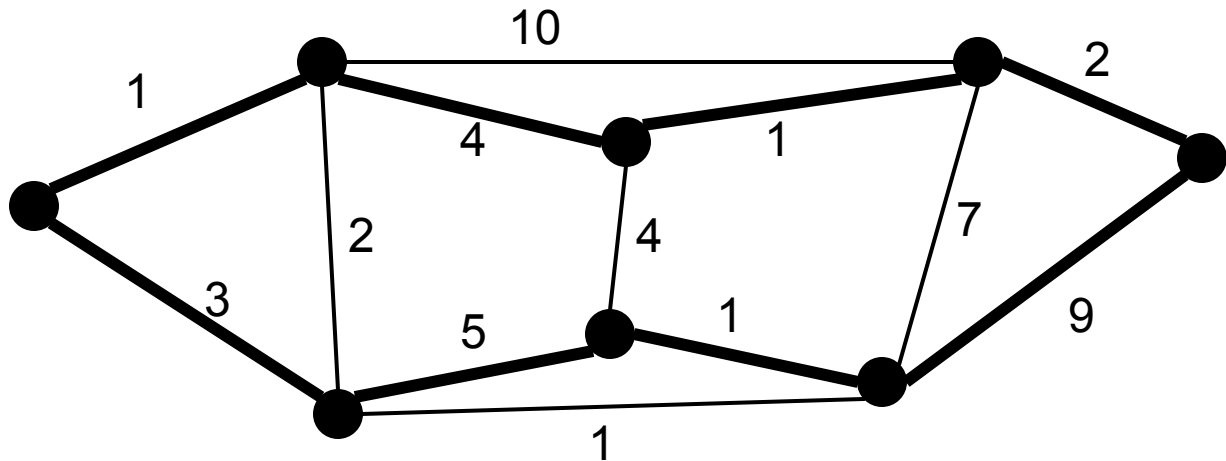


CAMHAM

- Lembra-se do algoritmo exponencial para CAM?
 - Ele fazia uma busca por todos os caminhos que passam por no máximo todos os nós de G . Cada um é um caminho potencial. (passo 1)
 - Para cada caminho potencial, basta testar se o mesmo é um caminho hamiltoniano. (passo 2)
 - Ou seja, o passo 1 leva tempo exponencial, mas o passo 2 leva tempo polinomial
- Bom, ainda assim, são necessários os 2 passos, o que resulta em tempo total exponencial
 - Mas já descobrimos algo.... vejamos outro exemplo

TSP – Traveling Salesman Problem

- O problema do caixeiro-viajante consiste em, dado um grafo não-direcionado G , onde cada aresta possui um peso inteiro, e um limite de peso W , encontrar um circuito hamiltoniano de peso total no máximo W



$W=28$

TSP

- Não foi encontrada ainda uma solução* que execute em tempo menor do que exponencial, mas:
 - Uma vez que se gaste tempo exponencial para encontrar uma solução (passo 1)
 - Pode-se, em tempo polinomial, verificar que a mesma é válida (passo 2)
- Novamente: o tempo total é exponencial
 - Mas existe um passo polinomial envolvido

* Existem soluções aproximadas bastante satisfatórias

Verificabilidade polinomial

- Alguns problemas, muito embora não conheçamos uma forma rápida (polinomial) de determinar uma solução
 - Uma vez descoberto, é possível verificar que a solução é válida
- Em CAMHAM, encontrar um caminho hamiltoniano é difícil, mas verificar que ele existe mesmo e é real, é fácil
- No TSP, encontrar um circuito hamiltoniano cujo peso total não ultrapasse um limite é difícil, mas verificar que ele existe mesmo e é real, é fácil

Verificador

- Um verificador para uma linguagem A é um algoritmo B , onde
 - $A = \{w \mid V \text{ aceita } (w,c) \text{ para alguma cadeia } c\}$
- Ou seja, dada uma cadeia w , e alguma informação adicional c , o verificador consegue determinar que w é membro de A
 - Essa informação extra c é chamada certificado, ou prova, da pertinência a A
 - Ex:
 - No CAMHAM, c é o caminho hamiltoniano de s a t
 - No TSP, c é o circuito hamiltoniano com peso máximo W

Verificador

- Um verificador de tempo polinomial executa em tempo polinomial no comprimento de w
- Ou seja, caso exista um verificador em tempo polinomial V para uma linguagem A
 - A é polinomialmente verificável

A classe NP

- Chegamos a outra definição vital de uma classe de linguagens

NP é a classe de linguagens que têm verificadores de tempo polinomial

- Obviamente, todo problema em P também está em NP ($P \subseteq NP$)
 - Mas veremos mais sobre isso daqui a pouco

A classe NP

- NP = Nondeterministic Polynomial time
 - Ou: tempo polinomial não-determinístico
- Existe uma caracterização alternativa que usa máquinas de Turing como base

NP é a classe de linguagens que são decidíveis em tempo polinomial sobre uma máquina de Turing não-determinística de uma fita

- Novamente, é óbvio que todo problema em P também está em NP ($P \subseteq NP$)
 - Mas calma

A classe NP

- É fácil imaginar a relação entre a definição com base em verificadores e a definição com base em MTs
 - V é um algoritmo, portanto existe uma MT M_v que o implementa
 - Se V é polinomial, a M_v também é
- M_v aceita entrada (w, c) em tempo polinomial
 - Mas precisamos construir uma NTM M que aceita w (sem c) em tempo polinomial
 - Construiremos M de forma a executar 3 passos:
 - M primeiro “adivinha” c , usando seu poder não-determinístico
 - Em seguida, M simula M_v
 - Se M_v aceita, M aceita. Se M_v rejeita, M rejeita

A classe NP

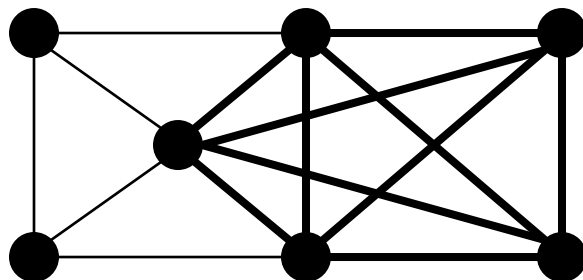
- O passo 1 é polinomial com relação ao tamanho de c (na verdade, basta “adivinhar” $|c|$ símbolos, portanto o passo 1 leva um tempo constante)
 - Adicionalmente c tem tamanho polinomial em relação à entrada w
 - Pois de outra forma, V não seria polinomial
 - Portanto o passo 1 é polinomial em relação à entrada w
- O passo 2 é polinomial devido à própria definição de V
- O passo 3 executa obviamente em tempo constante (na verdade, apenas um movimento é necessário)
- Portanto, M executa em tempo polinomial

Exemplos de problemas em NP

- Veremos agora alguns exemplos de problemas em NP
- Na verdade, os problemas anteriores, CAM, PRIM-ES, CAMHAM e TSP todos são exemplos de problemas em NP
 - Sabemos que CAM e PRIM-ES estão em P também
 - Mas para outros, não sabemos se estão em P (CAMHAM e TSP)
 - A seguir, apresentaremos alguns problemas que estão em NP, e aparentemente não estão em P

CLIQUE

- Um clique em um grafo não-direcionado é um subgrafo no qual todo par de nós está conectado por uma aresta. Um k-clique é um clique que contém k nós.
- $\text{CLIQUE} = \{(G,k) \mid G \text{ é um grafo não-direcionado com um k-clique}\}$



Um grafo
com um
5-clique

CLIQUE

- CLIQUE está em NP
- Prova:
- O clique é o certificado c
- Um verificador V para CLIQUE:
 - Teste se c é um conjunto de k nós em G
 - Teste se G contém todas as arestas conectando nós em c
 - Se ambos os testes retornam positivo, aceite. Caso contrário, rejeite.
- Prova alternativa
- Uma NTM polinomial que decida CLIQUE:
 - Mesmo procedimento acima, mas utilizando um passo não-determinístico que “adivinha” c

SOMA-SUBC

- Temos uma coleção de números x_1, x_2, \dots, x_k e um número-alvo t . Desejamos determinar se a coleção contém uma subcoleção que soma t .
- $\text{SOMA-SUBC} = \{(S, t) \mid S = \{x_1, x_2, \dots, x_k\} \text{ e para algum } \{y_1, y_2, \dots, y_l\} \subseteq S, \text{ temos } \sum y_i = t\}$
- $S = \{4, 11, 16, 21, 27\}$
- $t = 25$
- (S, t) pertence a SOMA-SUBC
 - Pois $4 + 21 = 25$

SOMA-SUBC

- SOMA-SUBC está em NP
- Prova
- O subconjunto é o certificado c
- Um verificador V para SOMA-SUBC:
 - Teste se c é uma coleção de números que somam t
 - Teste se S contém todos os números em c
 - Se ambos os testes retornam positivo, aceite. Caso contrário, rejeite.
- Prova alternativa
- Uma NTM polinomial que decide SOMA-SUBC
 - Mesmo procedimento acima, mas usando um passo não-determinístico que encontra c

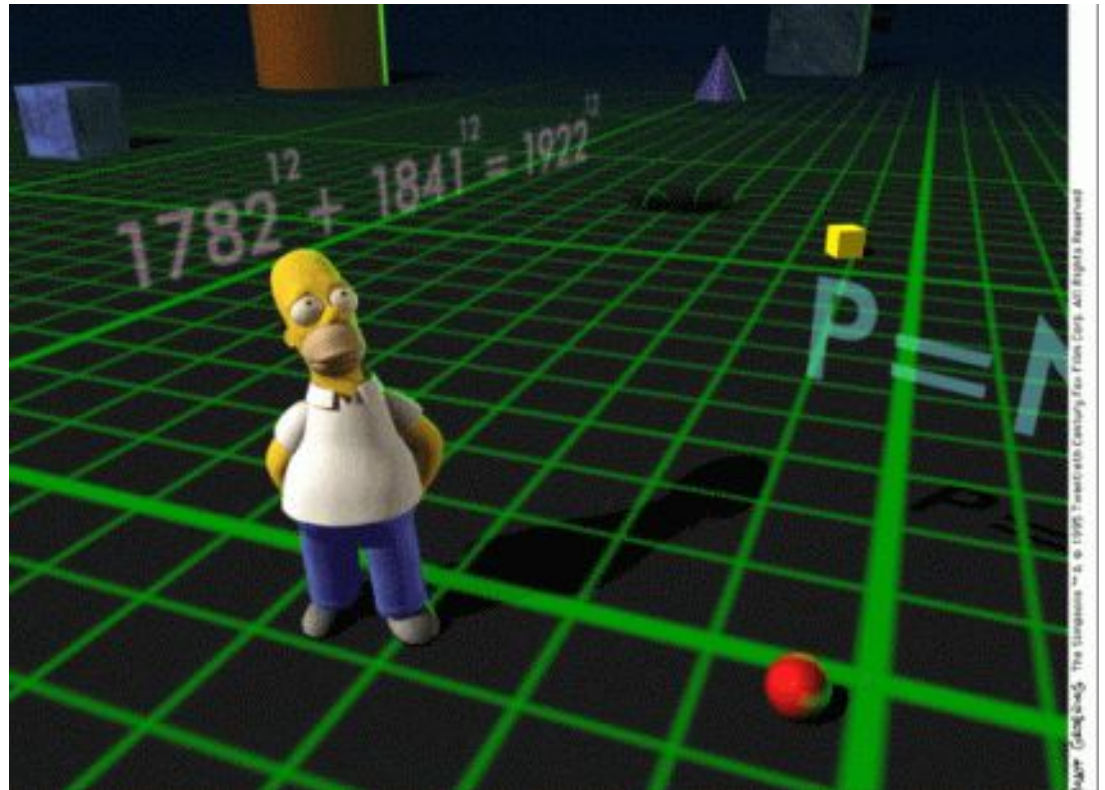
P versus NP

Curiosidades

- Episódio “Casa da árvore dos horrores VI” dos Simpsons
 - Exibido em outubro de 1995
- Três histórias curtas
 - Uma onde os personagens gigantes de outdoors ganham vida e atacam Springfield
 - Outra onde o Willie ataca de Freddy Krueger
 - Homer³: as irmãs da Marge vão visitar a família, Homer se esconde atrás do armário, e acaba encontrando um portal para a terceira dimensão

Homer³

- $1782^{12} + 1841^{12} = 1922^{12}$
 - “Contra-prova” do último teorema de Fermat
- $P=NP$

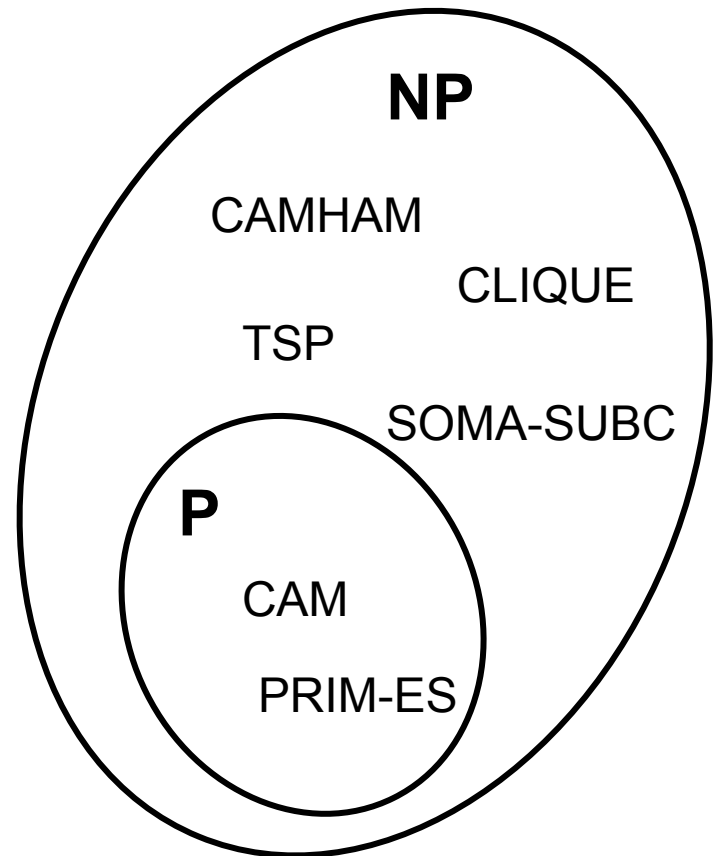


P versus NP

- P = classe das linguagens para as quais pertinência pode ser DECIDIDA rapidamente
- NP = classe das linguagens para as quais pertinência pode ser VERIFICADA rapidamente
- Existem muitos exemplos de problemas em P
 - CAM, PRIM-ES
- Existem muitos exemplos de problemas em NP, que aparentemente não estão em P
 - CAMHAM, CLIQUE, TSP, SOMA-SUBC

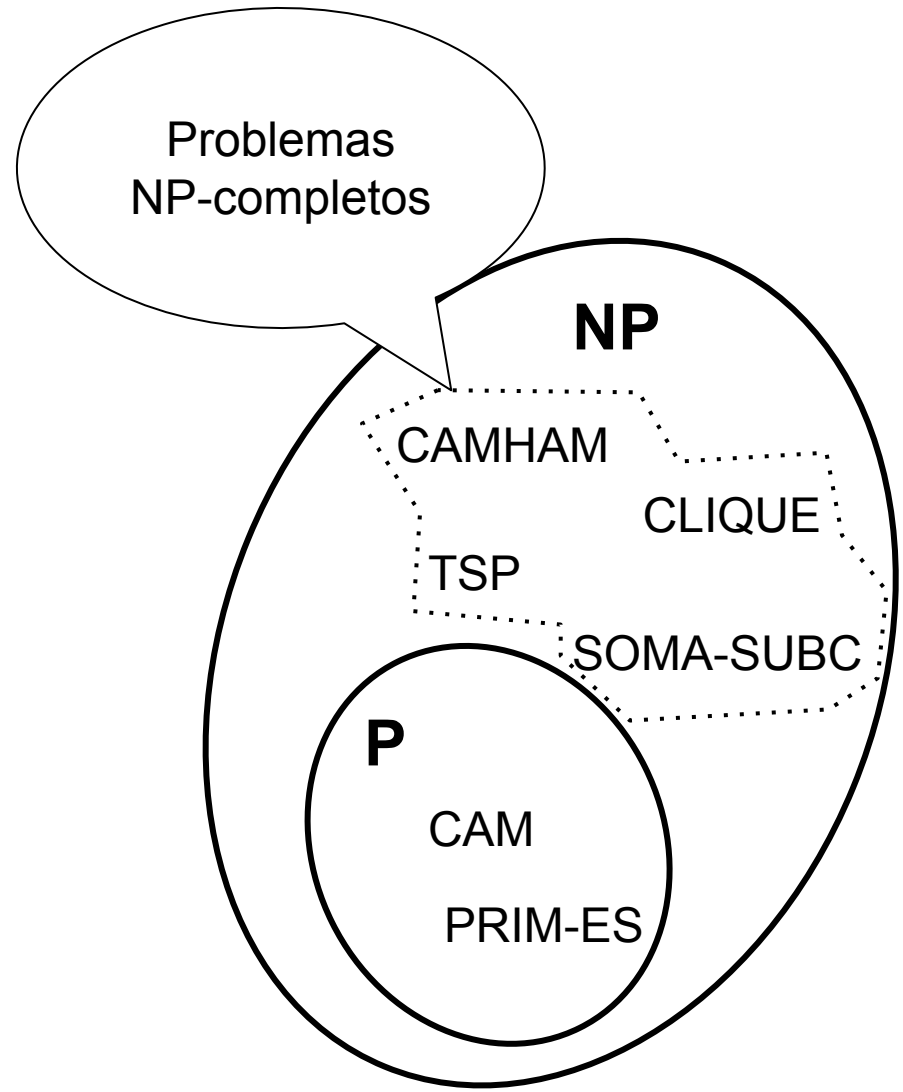
P ≠ NP?

- Aparentemente $P \neq NP$
- Porque “aparentemente”?
 - Porque muitas pessoas já tentaram resolver esses problemas em tempo menor que exponencial, e nunca conseguiram.
- Ou seja, apesar de provável, ninguém ainda conseguiu provar a existência de uma única linguagem que está em NP, mas não está em P



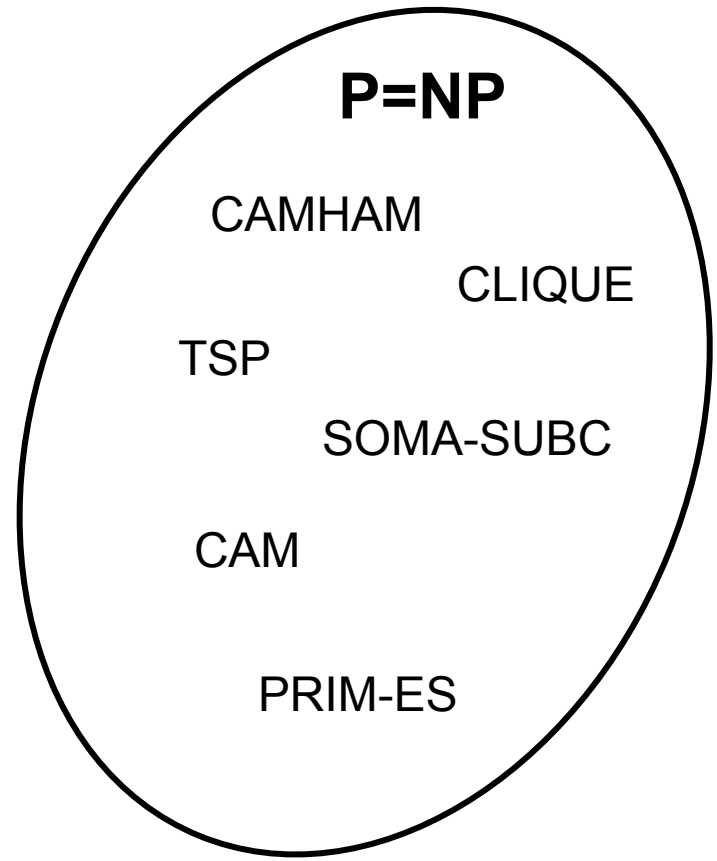
P ≠ NP?

- Tais problemas são chamados de NP-completos
- Uma classe especial de problemas
 - Encapsula justamente o ponto de dúvida
- Se $P \neq NP$
 - Todo problema NP-completo compartilha essa propriedade de complexidade
 - Ou seja, é impossível decidi-lo em tempo polinomial



P = NP?

- Mas pode ser que $P = NP$!
 - Talvez exista algum princípio matemático desconhecido
- Se $P=NP$, qualquer problema polinomialmente verificável seria polinomialmente decidível



P versus NP

- Como provar que $P = NP$?
 - Basta encontrar um único algoritmo de tempo polinomial que decide um único problema NP-completo
 - Com isso, todos problemas NP-completos poderiam ser decididos também, pois todos são relacionados entre si
 - Ou: basta encontrar uma forma mais eficiente de simular uma NTM em uma DTM
 - Até agora, o melhor método conhecido leva tempo exponencial

P versus NP

- Como provar que $P \neq NP$?
 - É necessário provar que não existe algoritmo rápido para substituir a força bruta
 - Ou provar que o método conhecido, de tempo exponencial, para simular uma NTM em uma DTM, é o melhor possível

E se $P=NP$?

- Primeiro: a prova iria levar eventualmente a métodos eficientes para resolver TODOS problemas NP-completos
 - E são muitos
- Criptografia, por exemplo, depende da dificuldade em resolver problemas NP-completos
 - Primalidade de um número e algoritmos de chave pública
 - Consequências “negativas” = teremos que pensar em novas maneiras para proteger informações

E se $P=NP$?

- Logística: problema do caixeiro-viajante
 - Consequências “positivas” = otimizar a utilização de meios de transporte
- Biologia: predição de estruturas protéicas tridimensionais
 - Consequências “positivas” = avanços na medicina e farmácia
- Matemática: prova formal de qualquer teorema
 - Exemplo: último teorema de Fermat levou 3 séculos para ser provado

E se $P \neq NP$?

- As consequências seriam menos catastróficas
 - Obs: aqui catástrofe = mudanças rápidas, para o bem e/ou para o mal
- Na verdade, a maioria trabalha com base nessa hipótese
 - Cada nova tentativa frustrada de provar que $P=NP$ aumenta a confiança de que $P \neq NP$
 - Décadas de tentativas são quase que uma prova “estatística” de que $P \neq NP$!
 - Como consequência muitos pesquisam soluções eficientes parciais, ou aproximadas, ao invés de tentar encontrar uma solução eficiente que seja exata
- Mas se alguém achasse a prova formal, teríamos a certeza de que estamos no caminho certo

Fim

Aula 28 - Problemas P e NP