

Análise de Diferentes Técnicas de Amostragem Para Coleta de Dados na Web 2.0

Thalisson Luiz Vidal de Oliveira
Fabrício Benevenuto

UFOP - Universidade Federal de Ouro Preto

14 de julho de 2011

- 1 Introdução
- 2 Objetivo
- 3 Trabalhos Relacionados
- 4 Metodologia
- 5 Experimentos
- 6 Análise
- 7 Complexidade
- 8 Trabalhos Futuros
- 9 Conclusão

Com constante o crescimento do volume de informações disponível na Web, existe também o crescimento da necessidade de se conhecer e estudar essas informações.

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-first search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

- Forma de se coletar dados na Web.
 - Busca em Largura
 - BFS (breadth-frist search)
 - Snowball
 - Busca em Profundidade
 - Forest Fire
 - DFS (Depth-First Search)
 - Busca por grau
 - Greedy
 - Lottery

Investigar o impacto de diferentes estratégias de amostragem na coleta, comparando a quantidade de vértices e arestas encontradas com o BFS e com o Snowball em diferentes níveis.

- **Caminho Mínimo**
- Distância entre pontos
- Estudos comportamentais

- Caminho Mínimo
- Distância entre pontos
- Estudos comportamentais

- Caminho Mínimo
- Distância entre pontos
- Estudos comportamentais

Para esse trabalho pretendemos implementar o método de busca em largura, utilizando redes P2P disponibilizados gratuitamente pela Stanford University por meio da biblioteca SNAP.

- O **BFS (breadth-frist search)** é um algoritmo que percorre por completo um grafo em largura, a partir de um determinado nó dado como raiz, expandindo e examinando todos os nó vizinhos até que todos os nós possíveis sejam coletados.
- O **Snowball** possui um processo de coleta similar ao BFS realizando uma busca em largura, a diferença entre ambos esta na condição de parada que para o Snowball ocorre quando todos os nós descobertos na raiz são coletados.

- O **BFS (breadth-frist search)** é um algoritmo que percorre por completo um grafo em largura, a partir de um determinado nó dado como raiz, expandindo e examinando todos os nó vizinhos até que todos os nós possíveis sejam coletados.
- O **Snowball** possui um processo de coleta similar ao BFS realizando uma busca em largura, a diferença entre ambos esta na condição de parada que para o Snowball ocorre quando todos os nós descobertos na raiz são coletados.

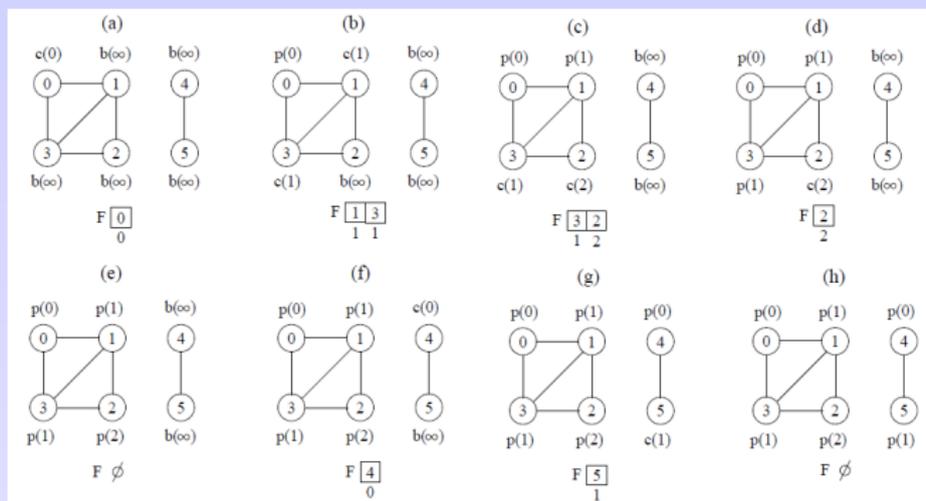


Figura: Exemplo de Busca em Largura [?]

- Quantidade de vértices encontrados.
- Quantidade de arestas encontradas.
- Caminho percorrido.

Experimentos

- Quantidade de vértices encontrados.
- Quantidade de arestas encontradas.
- Caminho percorrido.

- Quantidade de vértices encontrados.
- Quantidade de arestas encontradas.
- Caminho percorrido.

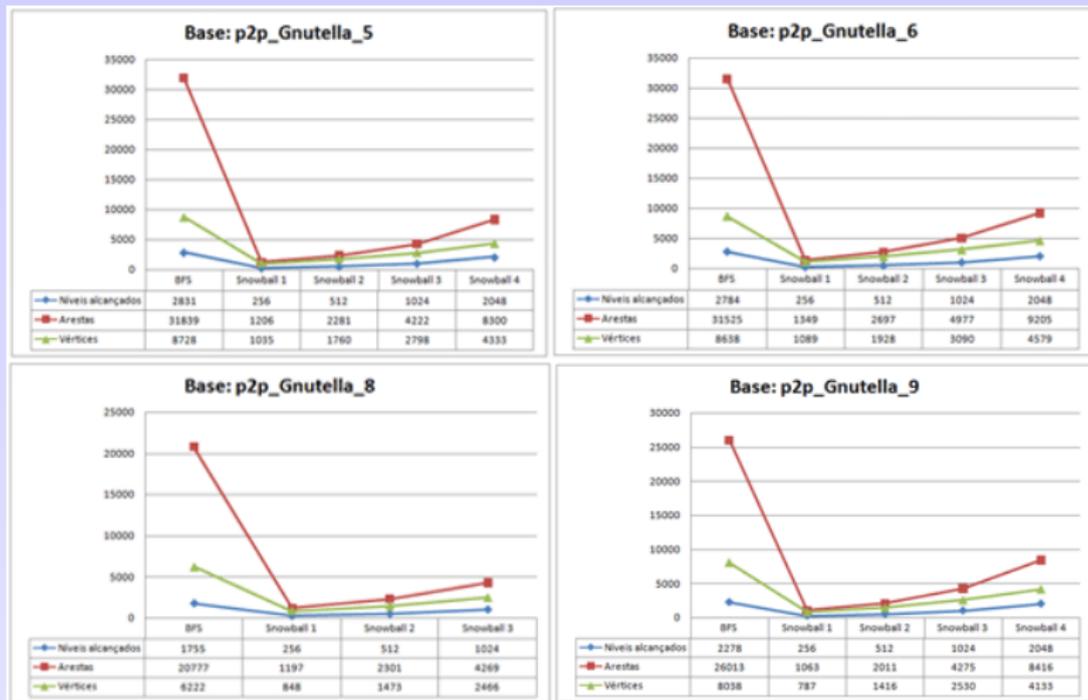


Figura: Resultado obtidos

- **BFS - Breadth-frist search**

- Complexidade de tempo: $O(|n| + |V^2|)$
- Complexidade de espaço: $O(m + 3V)$

- **Snowball**

- Complexidade de tempo: $O(|n| + |V * Corte|)$
- Complexidade de espaço: $O(m + 3V)$

- BFS - Breadth-frist search
 - Complexidade de tempo: $O(|n| + |V^2|)$
 - Complexidade de espaço: $O(m + 3V)$
- Snowball
 - Complexidade de tempo: $O(|n| + |V * Corte|)$
 - Complexidade de espaço: $O(m + 3V)$

- BFS - Breadth-frist search
 - Complexidade de tempo: $O(|n| + |V^2|)$
 - Complexidade de espaço: $O(m + 3V)$
- Snowball
 - Complexidade de tempo: $O(|n| + |V * Corte|)$
 - Complexidade de espaço: $O(m + 3V)$

- BFS - Breadth-frist search
 - Complexidade de tempo: $O(|n| + |V^2|)$
 - Complexidade de espaço: $O(m + 3V)$
- Snowball
 - Complexidade de tempo: $O(|n| + |V * Corte|)$
 - Complexidade de espaço: $O(m + 3V)$

- BFS - Breadth-frist search
 - Complexidade de tempo: $O(|n| + |V^2|)$
 - Complexidade de espaço: $O(m + 3V)$
- Snowball
 - Complexidade de tempo: $O(|n| + |V * Corte|)$
 - Complexidade de espaço: $O(m + 3V)$

- BFS - Breadth-frist search
 - Complexidade de tempo: $O(|n| + |V^2|)$
 - Complexidade de espaço: $O(m + 3V)$
- Snowball
 - Complexidade de tempo: $O(|n| + |V * Corte|)$
 - Complexidade de espaço: $O(m + 3V)$

Complexidade

```
for  $i = 0$  to  $i = V$  do  
  for  $j = 0$  to  $j = V$  do  
    if  $grafo_{ij} == 1$  then  
       $grafo_{ij} \leftarrow 2$   
       $QtdAresta ++$ ;  
      if  $visit_j \neq j$  then  
         $visit_j \leftarrow j$ ;  
         $fila_x \leftarrow j$ ;  
         $x ++$ ;  
         $no ++$ ;  
         $n ++$ ;  
      end if  
    end if  
  end for  
   $nivel_{i+1} \leftarrow no$   
   $no \leftarrow 0$   
end for
```

- Melhorar a complexidade dos algoritmos implementados.
- Testar em bases de redes diferentes.
- Estudar novos métodos de busca.

- Melhorar a complexidade dos algoritmos implementados.
- Testar em bases de redes diferentes.
- Estudar novos métodos de busca.

- Melhorar a complexidade dos algoritmos implementados.
- Testar em bases de redes diferentes.
- Estudar novos métodos de busca.

O resultado alcançado podem ser tendencioso dependendo do método utilizado e da abordagem escolhida.