

Aplicação de conceitos analíticos em grafos construídos a partir da relação entre curtidas de páginas do Facebook.

Kevin Regis Perondi¹, Matheus Sapia Guerra², Wesley Franco Ferreira³

¹Universidade Tecnológica Federal do Paraná - Campo Mourão (UTFPR)
Caixa Postal 271 – 87301-899 – Campo Mourão – PR – Brazil

kevinpregis@hotmail.com, guerramatheus2@gmail.com, wesley2ff@gmail.com

Abstract. *This article aims to apply the analyzes in a graph constructed from a list of tanneries between pages. An approach used in this article is applicable to an analysis to find the influential pages, and is measured or degree of each vertex.*

Resumo. *Este artigo tem por objetivo aplicar possíveis análises em um grafo construído a partir de uma relação de curtidas entre páginas. A abordagem utilizada neste artigo é aplicar uma análise para encontrar possíveis páginas influentes, isto é feito medindo o grau de cada vértice.*

1. Introdução

A técnica de usar grafos para resolver problemas do cotidiano vem sendo utilizada a séculos, é possível modelar quase tudo que existe no mundo em um grafo, como por exemplo, o caminho mínimo entre duas cidades, o gasto mínimo para se construir uma infraestrutura de rede, conexão de amizade nas redes sociais, entre outros. Com a ascensão das redes sociais, a análise de uma possível conexão entre pessoas, páginas ou podem render informações relevantes.

Neste Artigo é modelado um grafo a partir de uma relação entre curtidas de páginas do Facebook, ou seja, cada página é traduzida para um vértice e as arestas representam a relação de curtidas entre estas páginas. Com um conjunto de vértices e arestas obtemos um grafo, e assim é possível aplicar algumas operações analíticas no mesmo. Neste experimento foi aplicado uma métrica que conta o grau de saída e entrada de cada vértice, ou seja, essa métrica conta a relação de curtidas entre páginas. Com isso é possível obter conclusões a partir dos dados calculados.

2. Fundamentação

Este capítulo tem por finalidade discorrer sobre ferramentas, técnicas e conceitos importantes para a realização do experimento. A subseção mostra como uma rede social é de fato construída como um grafo. A seção blá descreve conceitos sobre grafos, assim como suas subseções descreve sobre vértices, arestas e técnicas que podem ser aplicadas em um determinado grafo. A última seção tem por objetivo apresentar a ferramenta utilizada para implementar o experimento, tal como a biblioteca utilizada.

2.1. Redes sociais

Redes sociais são estruturas sociais compostas parte por pessoas parte por organizações contendo uma ou várias relações entre si. Uma das características mais fundamentais

sobre rede sociais é sua abertura para relações não hierárquicas entre os participantes, mas mesmo sendo fundamental a abertura entre participantes, as conexões sociais são utilizando identidade, geralmente com nome e sobrenome ou nome da organização. Sendo assim é possível dizer que uma rede social tem propriedades para observar padrões de conexão entre os elementos, sejam eles pessoas ou páginas.[e José Santos 2015] A Figura 1 permite observar essas conexões descritas com mais intuitividade.

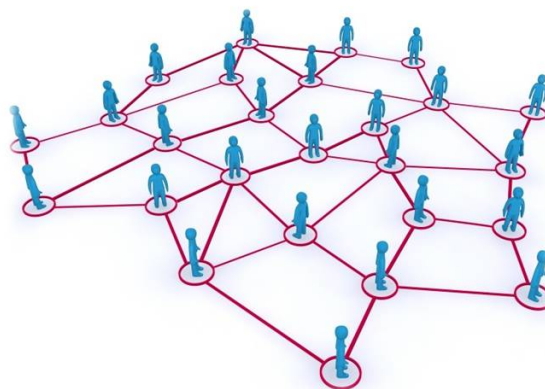


Figura 1. Exemplo de como uma rede social pode ser denotada com um grafo.

A rede social selecionada para utilização deste trabalho será o Facebook. O Facebook foi considerado a maior rede social com 1,15 bilhão de usuários únicos. Seu criador e fundador é Mark Zuckerberg. Para utilizar o Facebook o usuário deve declarar ter pelo menos a idade mínima de 13 anos. Atualmente as redes sociais estão praticamente em tudo no dia-a-dia das pessoas, seja postando uma foto para familiares até marcando ou até mesmo realizando reuniões de grandes corporações, pensando nisso será detalhado neste relatório o desenvolvimento de uma aplicação real para a disciplina de grafos.

2.2. Grafo

Um Grafo é, basicamente, um conjunto de vértices ' V ' e um conjunto de arestas ' A ', na qual este último tem por finalidade conectar os vértices entre si. O grafo pode ser dirigido (conhecido também como dígrafo) onde suas arestas possui direção, ou seja, uma aresta tem um vértice de origem e um de destino, ou um grafo pode ser não dirigido, onde suas arestas implicitamente são bidirecionais, pode-se dizer que seria análogo a um grafo dirigido onde exista uma aresta com um vértice de origem $V1$ com destino ao vértice $V2$ e $V2$ tenha uma aresta com vértice de origem $V2$ com destino ao vértice $V1$. O interessante em um grafo é que muitas coisas no cotidiano pode ser representadas por um grafo de várias formas diferentes.[Bondy 1976]. A Figura 2 ilustra um dígrafo, onde as arestas são direcionais.

Os vértices apresentados obrigatoriamente deve ter um rótulo, que deve ser único em um grafo, outro atributo que um vértice pode ter é cores, muito utilizado em buscas (largura, profundidade por exemplo), outro atributo importante em um vértice é o predecessor que é o "pai" de um determinado vértice.

As arestas, pode ter um atributo que é o peso. O peso da aresta é importante pois pode ajudar a abstrair muitas coisas da vida real, tal como o custo de ir de uma cidade

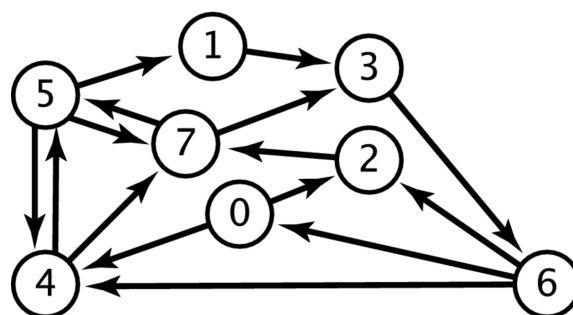


Figura 2. Exemplo de um Dígrafo.

até outra, sendo que o custo pode ser distância, condições da via, pedágio, entre outros. Sendo assim o custo, ou peso de uma aresta pode ser definido de acordo com o problema a ser modelado.

2.2.1. Operações Analíticas em Grafos

Um grafo pode ter várias operações analíticas, pode-se citar por exemplo, busca em largura e busca em profundidade, árvore geradora mínima, caminho mínimo, entre outras. Cada uma dessas operações analíticas pode ser usada para resolver ou correlacionar itens, inclusive do mundo real. Neste experimento foi utilizada uma métrica que consiste em contar o grau de entrada e saída de cada vértice.

2.3. NetBeans

O NetBeans é uma ambiente de desenvolvimento integrado, muito utilizada por programadores pois oferece ferramentas necessárias para criar aplicações de diversos tipos. O grande diferencial do NetBeans é que o mesmo possui uma vasta quantidade de API, bibliotecas e uma documentação completa em várias linguagens diferentes[NET 2017].

2.3.1. Biblioteca Restfb

O RestFb é uma, é uma biblioteca *open source* que realiza operações da API do Facebook. É escrita em Java, tem suporte para projetos em Android, e não depende de outras bibliotecas para ser utilizada. Neste experimento foi necessário somente instalar a biblioteca restfb e já estava pronto o ambiente.

3. Metodologia

Para o cumprimento deste trabalho será realizado uma coleta de dados a partir de uma página na rede social Facebook, os dados verificados serão de curtidas que esta página recebe na rede social. Além disso, para manter uma estrutura com uma baixa desigualdade de curtidas em cada página, um valor limite de 25 curtidas será estabelecido.

A obtenção do grafo da rede será baseado em uma coleta de acordo com um raio de busca, cujo o funcionamento é determinado da seguinte maneira:

1. Seleção de uma página semente;

2. Coleta das páginas curtidas pela página semente;
3. A partir do valor do raio, a operação de coleta se estende sucessivamente.

Para melhor compreensão do funcionamento do algoritmo de coleta, a Figura abaixo representa o funcionamento deste coletor.

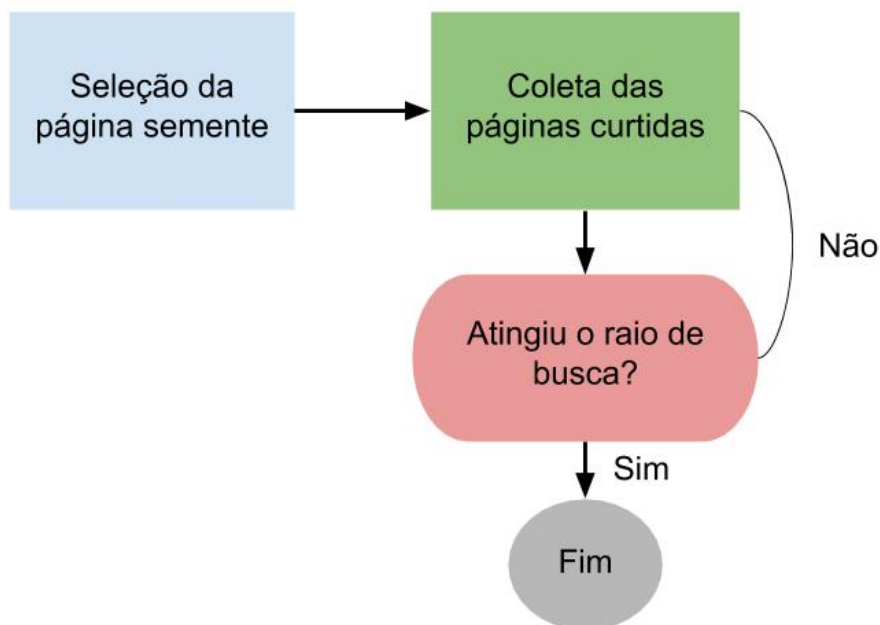


Figura 3. Funcionamento do coletor das curtidas

A estrutura do grafo é basicamente compreendida da seguinte maneira, cada página coletada será representada por um vértice e as arestas serão as curtidas estabelecida entre as páginas. Para a visualização do grafo desta rede, utilizaremos a ferramenta *online* Webgraphviz¹ que possibilita a visualização por meio de uma arquivo de entrada contendo especificações sobre as arestas do grafo.

Por fim, para obter uma análise desta rede será aplicado a medida do grau em cada vértice com a finalidade de identificar qual a página com a maior interação nessa rede.

4. Resultados e Discussões

Nesta Seção serão apresentados os métodos implementados juntamente com sua complexidade e os resultados obtidos a partir da aplicação da medida do grau na rede das páginas analisadas.

4.1. Implementação e Complexidade

Iniciando uma discussão a respeito do coletor, podemos afirmar que seu crescimento será definido de acordo com o valor do raio. Caso o raio de coleta seja de 1, sabemos que a complexidade é estabelecida pelo limite de 25 curtidas por página. Para o valor do radio igual a 2 será coletado, além das curtidas da página semente, as curtidas das 25 páginas identificadas na primeira iteração. Assim, podemos definir que a complexidade da coleta

¹<http://www.webgraphviz.com/>

é definida pelo limite da coleta elevado ao valor do raio mais a soma da página semente, sendo $25^r + 1$.

Além da complexidade da coleta, há a complexidade das operações a serem realizadas pelo grafo. Todos métodos relacionados a inserção no grafo irá possuir a complexidade $O(n)$, essa complexidade deve-se ao fato da inserção ser segura, ou seja, é verificado a lista de vértices ou arestas para que não haja duplicação dos elementos. O método `checkEdges`, utilizado para estabelecer as arestas entre páginas coletadas, varre todos os vértices do grafo e para cada vértice verifica sua lista de quem este vértice curtiu, ainda, realiza a busca na lista de arestas antes de realizar essa nova inserção. Deste modo, o método `checkEdges` possui a complexidade $O(n^2)$, sendo ele o método mais custoso no grafo. Para o cálculo do grau, o método `calculaGrau` realiza a busca por todos os vértices do grafo realizando a chamada do método `arestasComEsteVertice`, que tem por finalidade contar a quantidade de arestas que cada vértice possui. Ambos os métodos possuem a complexidade $O(n)$.

4.2. Resultados

Utilizando a página “Skol” como semente, realizamos quatro coletas variando o valor do raio de busca e aplicando a medida do grau para identificar a página de maior influência dentro desta rede. A Tabela 4.2 apresenta os valores encontrados.

Tabela 1. Resultados da página “Skol”

Raio	Vértices	Arestas	Tempo de execução	Página influente
1	23	22	7s	“Skol”
2	316	347	10s	“Karon Conka”
3	3954	4458	1m e 30s	“Karon Conka”
4	40574	56244	22m e 24s	“Jamie Jones”

Desta forma, podemos observar um aumento gradativo no tempo de execução e na quantidade de vértices e arestas estabelecidos entre as coletas com diferentes raios. Também foi possível observar que a página “Karol Conka” permaneceu a mais fluente na rede mesmo com os valores de raio 2 e 3. E a página só perdeu seu posto de maior grau a partir do raio 4.

A visualização do grafo da rede foi elaborada por uma página cujo o número de curtidas fosse inferior a 25. Consultado o Facebook, encontramos a página “Kawasaki Ninja 300r” que apresentou um número total de 4 curtidas. A Figura 4 mostra o desenho da rede criado pela ferramenta Webgraphviz.

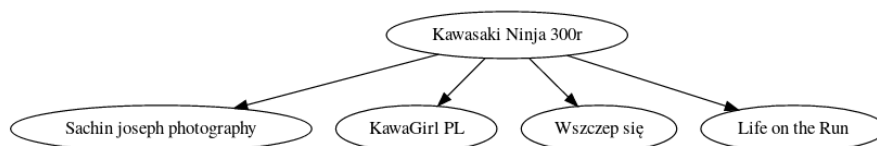


Figura 4. Estrutura do grafo para a página “Kawasaki Ninja 300r”

Grafos com maiores dimensões não é suportado pela ferramenta, impossibilitando a visualização completa pelo usuário.

5. Conclusão

Com o presente trabalho podemos concluir a importância do uso de grafos para analisar uma rede, possibilitando a extração de informações quantitativas presente na mesma por meio de métricas.

6. Recomendações para Trabalhos Futuros

Para trabalhos futuros, pode-se apontar:

- Implementação de motores de busca, tais como algoritmos de busca em largura podendo ser utilizado para encontrar um respectivo nó no grafo de curtidas, ou até mesmo percorrer o grafo verificando a proximidades dos nós tentando assim identificar a proximidade das "páginas". Podendo também identificar um nó central no qual o caminho para os demais vértices é o mais curto.
- Utilização da métrica de intermediação no qual verificar quantas vezes um vértice(uma página) é utilizada como ponte entre o caminho mais curto entre duas páginas. Começando com o calculo do caminho mais curto entre os dois vértices, para cada par de vértices com o caminho mais curto calculado determinar a fração de caminhos que passam no vértice em questão (vértice desejado para calcular quantas vezes se comportar como ponte) ao final de tudo somar as frações e determinar sua intermediação.
- Métricas Baseadas em Vizinhança Local algumas abordagens sugerem que dois vértices u e v possuem vizinhos (curtem a mesma página), estão propensos a gerar um laço.
 - Ligação preferencial propõe que princípio que páginas com maior número de conexões sociais são mais propensos a formar novos laços, o conjunto de ligação preferencial por si só diz que um vértice se conecta a outro vértice da rede com uma chance baseada em quem são os seus vizinhos.
 - Vizinhança em Comum identifica o número de vizinhos em comum entre os vértices. A ideia de vizinhos em comum é apenas uma representação de transitividade do grafo. Resumidamente se em um grafo gerado por uma página, caso o vértice V está ligado ao vértice U e o vértice W está ligado ao vértice V , então existe uma maior chance de o vértice W estar ligado ao vértice U , então se aumentar os vizinhos em comum entre V e U a chance entre W e U estar ligado também aumenta.

- Coeficiente de Jaccard mede a importância da vizinhança em comum, a medida define a chance de um vértice vizinho comum de um par de vértices u e v seja selecionado aleatoriamente, a partir da união de conjuntos u e v . Com isso quanto maior a incidência de vizinhos em comum maior será o coeficiente de Jaccard.
- Métricas Baseadas em Caminho mais Curto: O fato de as páginas curtidas de nossa página inicial existir a chance de se "curtirem" sugere que a distância entre dois vértices em uma rede social pode influenciar na formação de um laço entre eles. Dependendo da distância entre eles, por exemplo quanto mais perto a distância maior será a chance de ocorrer um laço. Para avaliação do mesmo podem ser usadas as métricas: Katz, Tempo de acerto, Tempo de Comutação e Rooted Pagerank.
- Katz; A medida Katz está relacionada ao conceito de caminho mais curto, e geralmente funciona com a predição de laços. Este atributo soma diretamente todos os caminhos que existem entre um par de vértices u e v . Para penalizar a contribuição de caminhos mais longos no cálculo de semelhança, amortece exponencialmente a contribuição de um caminho por um factor de Beta em que l representa o comprimento do percurso.
- Tempo de Acerto O tempo de acerto, ou *Hitting Time*, conta o número esperado de passos necessários para um passeio aleatório chegar do vértice u ao v . Um valor pequeno desta medida indica que os vértices são semelhantes uns aos outros, e tem uma chance maior de se ligarem no futuro. O benefício desta métrica é ser fácil de calcular por meio da realização de alguns passeios aleatórios.
- Tempo de Comutação O tempo de comutação, também conhecido com Comute Time, conta o número esperado de passos necessários para acessar o vértice u a partir de v , e no caminho de volta de v para u . Esta medida pode ser obtida com a Equação: $CT(u, v) = HT(u, v) + HT(v, u)$. Uma vez que o tempo de acerto não é uma medida simétrica para grafos não direcionados, ou seja, os valores de $HT(u, v)$ e $HT(v, u)$ não são iguais, o tempo de comutação pode ser usado.
- Rooted Pagerank esta medida é utilizada para o ranking das páginas, e tem relação intrínseca como o tempo de acerto, Além disso, pode ser usada como um recurso para a predição de laços sociais. No entanto uma vez que Pagerank é um atributo de um único vértice, necessita ser modificado de modo que represente uma similaridade entre um par de vértices x e y . Considerando uma probabilidade α fixa, uma página salta de uma para outra página aleatória com a probabilidade α e segue um hiperlink com probabilidade $1 - \alpha$. Para este passeio aleatório, a importância de uma página v é a soma esperada da importância de todas as páginas u que apontam para v . Na terminologia do passeio aleatório, pode-se substituir o termo importância pelo termo distribuição estacionária.

Referências

- (2017). Netbeans ide. Disponível em <https://netbeans.org/> Acessado em 05/12/2017.
- Bondy, J. A. (1976). *Graph Theory With Applications*. Elsevier Science Ltd., Oxford, UK, UK.
- e José Santos, V. S. (2015). As redes sociais digitais e sua influência na sociedade e educação contemporâneas. *HOLOS*, 6(0):307–328.