

Maximização de Influência

Luccas Guidio¹

¹Departamento de Ciência da Computação –
Universidade Federal de São João del Rei / UFSJ

²Núcleo de Educação a Distância / NEAD -
Universidade Federal de São João del Rei / UFSJ

{igorslbatista, guidioo}@hotmail.com

Abstract. *Influence Maximization is a very common problem nowadays, we see a lot of brands or businesses working hard to reach the highest number of people with their publicity/services spending as little as possible. The problem in question consists in finding the smallest group of individuals in a community that will be able to influence most of the other individuals in the same community. This article's objective is to explain, define and solve the problem of influence maximization*

Resumo. *Maximização de Influência é um problema muito comum nos dias atuais, vemos muitos tipos de marcas ou empresas se esforçando sempre para conseguir atingir o maior número de pessoas com suas propagandas/serviços gastando o menos possível. O problema em questão consiste em encontrar o menor grupo de indivíduos em uma comunidade que vai ser capaz de influenciar a maior quantidade de indivíduos nesta mesma comunidade. O objetivo do artigo é explicar, definir e resolver o problema de maximização de influência*

1. Introdução

O problema de entender como as pessoas influenciam umas às outras sempre existiu, grandes empresas ou marcas famosas tentam, diariamente, encontrar pessoas ou comunidades que possam divulgar ou propagar seus negócios da maneira mais barata possível. Uma loja x, por exemplo, acabou de lançar uma nova camiseta, porém não tem verba para a publicidade de seu novo produto, então decide dar de presente algumas unidades com o intuito de ativar a vontade de compra em outras pessoas. Como existe um número limitado de produtos grátis e nada garante que as pessoas terão vontade de comprar a peça de roupa, a loja precisa encontrar a menor

quantidade de pessoas que convenceriam o maior número de pessoas a testar seus produtos, então procura indivíduos que consigam *maximizar a influência* de sua marca. O exemplo anterior é um problema de maximização de influência, que é um problema NP-Difícil e definido como uma solução S com K indivíduos de maneira que a influência dos indivíduos de S seja maximizada em um grafo $G = (V, A)$ com os vértices representando pessoas ou comunidades que constituem o "mundo" e arestas A_{ij} quando um vértice V_i é capaz de influenciar um outro vértice V_j .

2. Conceitos Fundamentais

Um grafo $G = (V, A)$ possui vértices e arestas, como explicado anteriormente, os vértices são as pessoas ou grupos de pessoas que formam o seu conjunto Universo, e caso uma pessoa possua a possibilidade de influenciar outra, é criada uma aresta entre elas que sai da influenciadora e chega na influenciada. Estes grafos são conhecidos também por *Social Networks*, para realizar a maximização de influência, cada aresta e vértice possui uma propriedade chamada de ativação. Vértices que foram influenciados com sucesso são ativados, caso contrário permanecem inativos. Uma aresta é ativada quando um vértice *tenta* ativar o outro, todas as arestas são, inicialmente, indiferentes, ou do inglês *unflipped*, assim que um vértice V_i tenta ativar o V_j , a aresta entre ambos é alterada, porém apenas se a ativação é bem sucedida a aresta é ativada, caso contrário se torna inativa. Por conta da sua aleatoriedade, os resultados obtidos em dois testes iguais podem variar, claro que sempre existirá uma certa proximidade em resultados que partem de uma mesma configuração inicial, porém não serão exatamente iguais. Um outro termo importante é "sementes", sementes são os vértices ativados inicialmente no problema, para verificar sua propagação

3. Motivação

Foram estudados alguns trabalhos relacionados a maximização de influência para primeiramente entender o que seria, de fato, o problema, como tratar com cada tipo de *input* e ver algumas possíveis heurísticas para solucionar o problema. Apesar de terem sido lidos múltiplos artigos sobre o assunto, todos usavam *large-scale ou complex social networks* o que acabou não sendo muito útil uma vez que a aplicação desenvolvida, apesar de conseguir lidar com redes sociais de grande escala, não é otimizada para o tamanho de entrada que muitos dos artigos estavam utilizando. Foi utilizado o site pessoal de [Newman] para encontrar algumas entradas diferentes e re-

solver o algoritmo. Foram escolhidas três *social networks*, uma que se trata de 34 membros de um clube de karatê em uma universidade americana nos anos 70, uma outra que se trata de associações de golfinhos vivendo em um aquário na Nova Zelândia e por fim uma rede de livros suas conexões representam compras dos livros por um mesmo comprador

4. Metodologia

Foram desenvolvidas duas heurísticas para a resolução do problema de maximização de influência, resumidamente, a primeira consiste em encontrar os K melhores indivíduos em questão de influência no grafo e a segunda em calcular a influência de uma lista de indivíduos pré-definida.

Para ambas as heurísticas a entrada do programa pode ser uma matriz onde os vértices são os indivíduos em uma comunidade e as arestas $A(ij)$ representam a possibilidade de um indivíduo I influenciar um outro vértice J. A entrada do programa pode, também, ser uma lista de "arestas" onde cada linha do arquivo consiste em dois números, o primeiro representando de onde a aresta sai e o segundo onde ela termina. Em ambos os casos, é criada uma lista de "vértices" que são originados de uma classe com atributos representando se estão ativos ou não e também uma outra lista de vizinhos. Para cada elemento desta lista de vizinhos, existem quatro valores numéricos que correspondem ao estado da aresta que é representada. Se não existir uma aresta entre o vértice original e o I-ésimo elemento do vetor de vizinhos, é dado o valor 0, caso exista a aresta, ela é inicialmente dada como 1, porém este valor é alterado ao decorrer do algoritmo para representar se a aresta foi ativada com sucesso ou não (2 e 3 respectivamente)

Com a lista principal criada, o algoritmo decide, através da entrada, qual tipo de resolução vai tentar fazer, no caso de tentar procurar os K melhores indivíduos dessa lista, o algoritmo calcula a influência de cada elemento da lista "grafo" da seguinte maneira:

1 - Para cada elemento pertencente ao grafo original, adiciona ele à uma lista de solução, chamada de "listaSeed", caso ele já não esteja presente nela.

1.1 - Cada elemento pertencente a lista que estiver inativo, é ativado.

1.2 - Para cada vizinho relativo ao elemento da lista citada anteriormente, ativa a *aresta* entre ambos e tenta, então, ativar o vizinho em

questão, levando em consideração a probabilidade de ativação.

1.3 - Para cada um dos vizinhos ativados em resultado ao passo anterior, realiza o mesmo passo (1.2) tentando ativar seus vizinhos (vizinhos de vizinhos e assim em diante), dessa forma **todas** as arestas que saem de vértices ativos se encontrarão ativadas, ao final dessas repetições terá o resultado da influência dos elementos atuais da lista.

1.4 - Remove o elemento da lista escolhido em (1) da lista.

2 - Assim que todos os elementos do grafo original tiverem sido testados separadamente, é escolhido o de melhor resultado e adicionado novamente à lista de solução, porém, dessa vez ele não vai ser retirado desta.

3 - Volta para o passo (1) mas com uma lista inicial maior.

4 - Assim que a lista atingir K elementos, retornar a lista como solução.

Os passos de 1 e 2 são chamados de "calcula solução", que vai ser a função utilizada para a resolução da segunda heurística onde é chamada essa função uma vez para cada elemento da pré-definida, depois disso conta-se quantos vértices possuem o atributo "status" como "ativado" para obter o valor da influência dessa lista original.

Como a ativação de cada indivíduo é decidida por uma probabilidade, e portanto, possuir um certo grau de aleatoriedade, todos estes cálculos de solução são realizados 100 vezes e o resultado obtido é na verdade uma média de todos os resultados destas 100 repetições.

5. Resultados

A seguir alguns gráficos de resultados obtidos durante a execução do algoritmo para as três Networks utilizadas, o tempo de execução está em segundos e o tamanho da solução é o valor de K (K elementos que maximizam a influência) e a Influência Média é a média de quantos elementos foram ativados com as soluções obtidas após 100 repetições do algoritmo

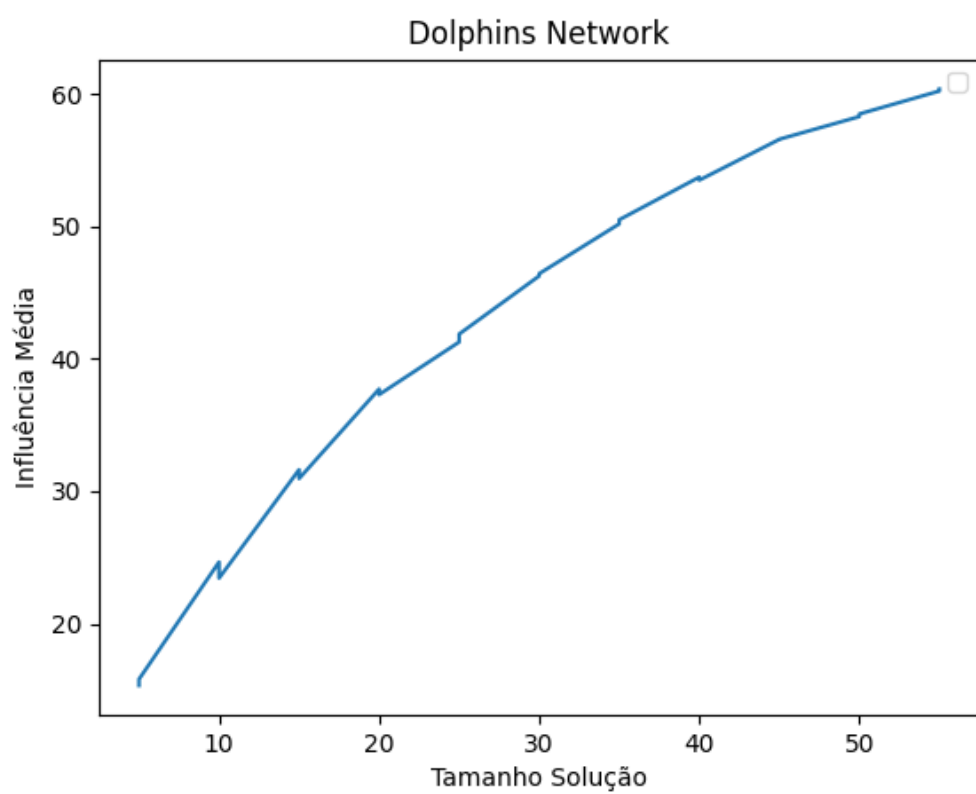


Figura 1.

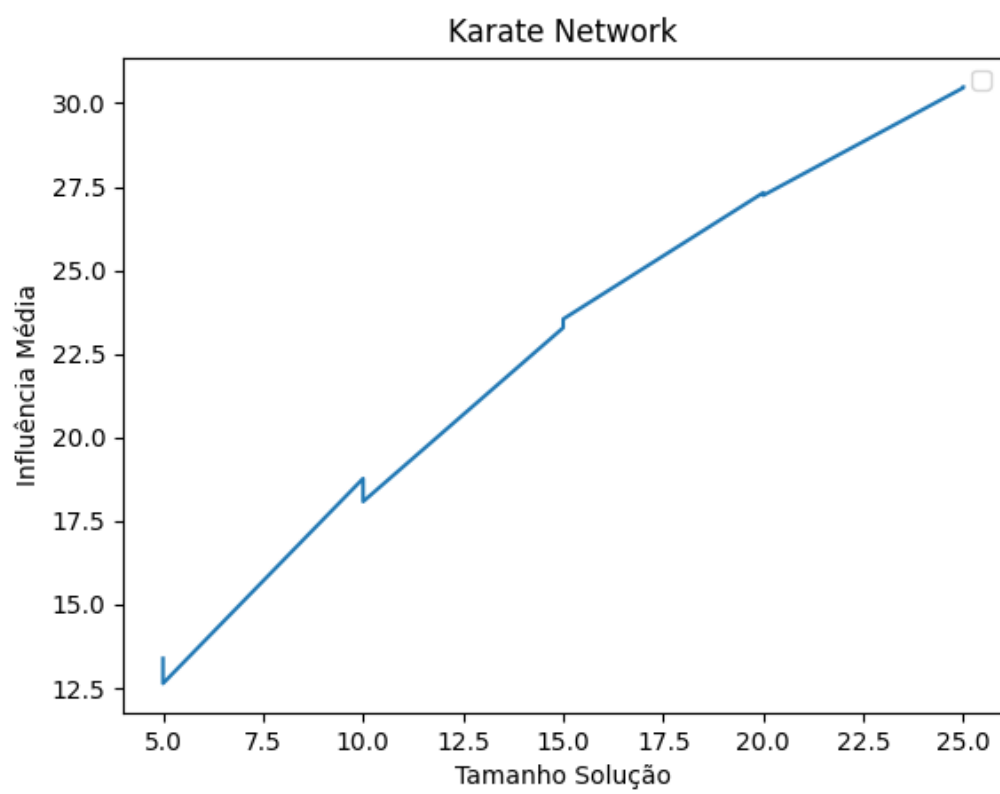


Figura 2.

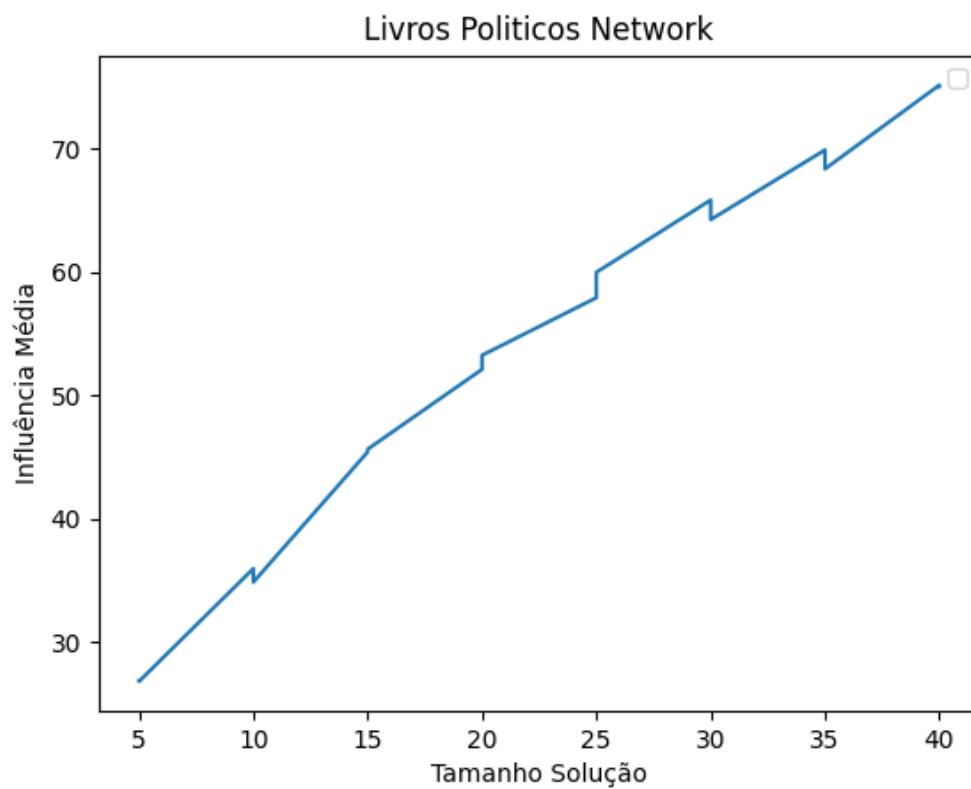


Figura 3.

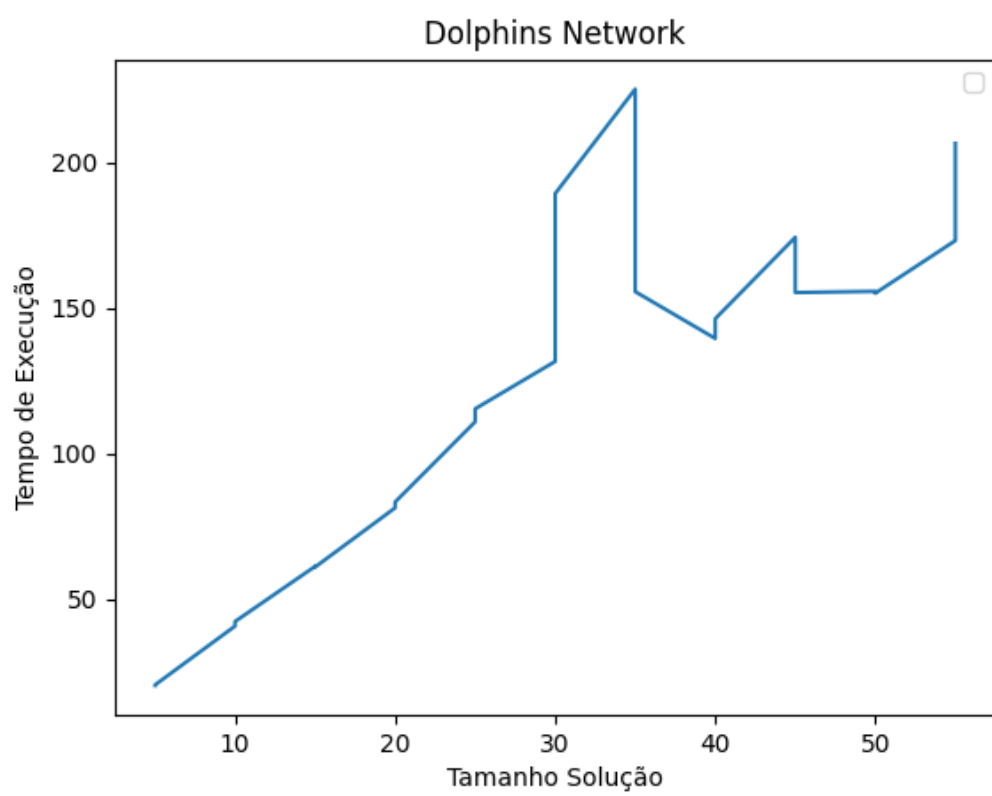


Figura 4.

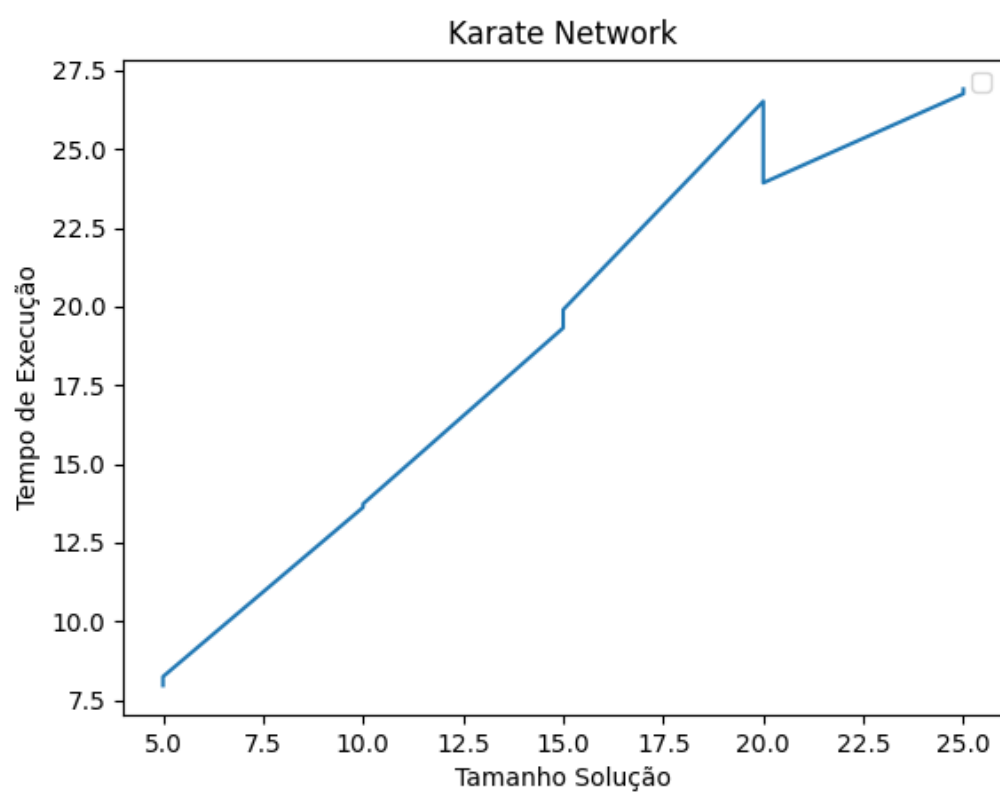


Figura 5.

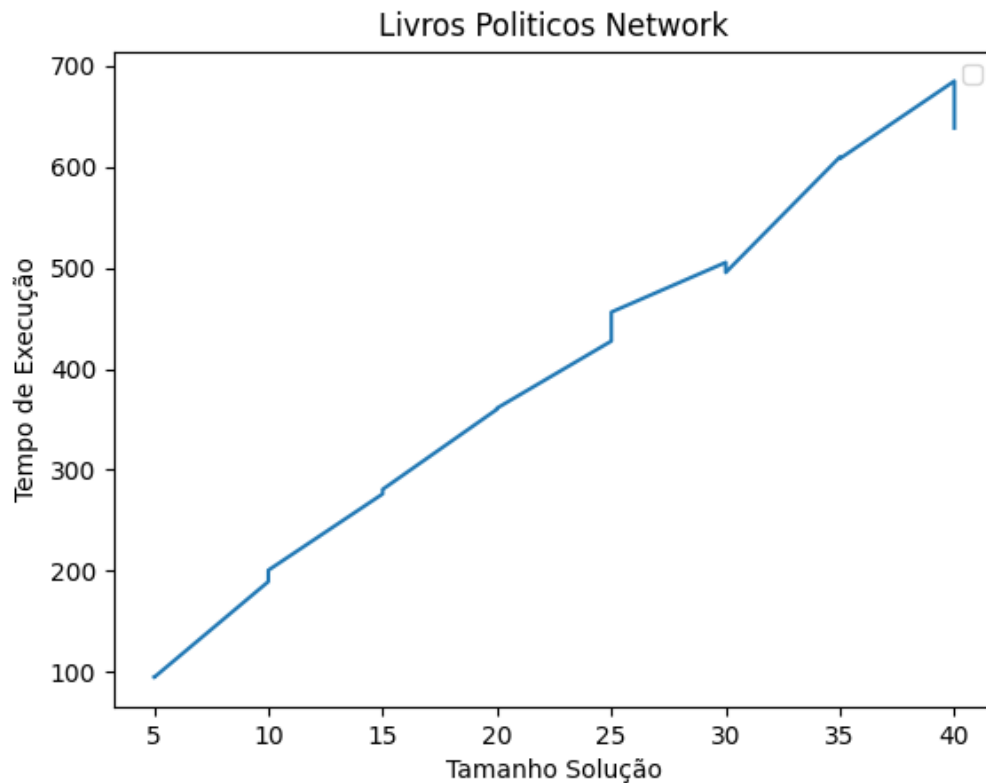


Figura 6.

6. Conclusão

Como o problema da maximização de influência é algo que encontramos no dia-a-dia a resolução dele é mais do que algo apenas acadêmico, é interessante entender como as pessoas são capazes de influenciar umas as outras e conseguir criar alguma representação computacional para lidar com esses dados é uma dificuldade enfrentada por muitos, porém é algo que está sendo muito utilizado, existem várias brincadeiras sobre o Facebook estar gravando suas conversas, onde você clica e vendendo este tipo de informação, essas informações de onde você navega quando está utilizando a internet são utilizadas para criar *Social Networks* entre pessoas que são amigas na rede, ajudando marcas e empresas famosas a indicar produtos que você tem mais probabilidade de comprar e recomendar para os seus amigos.

Com base no que foi apresentado, é possível entender como se descobre, a partir de grafos direcionados representando comunidades de pessoas, o maior *influencer* do grupo, e apesar dos exemplos não serem de

grande escala, a ideia continua a mesma, porém seria necessária uma otimização do algoritmo.

Referências

- Chen, W., Wang, Y., and Yang, S. (2016). Efficient influence maximization in social networks. https://www.microsoft.com/en-us/research/wp-content/uploads/2016/02/weic-kdd09_influence.pdf.
- Newman, M. personal website. <http://www-personal.umich.edu/~mejn/netdata/>.
- Xavier, C. R. (2015). Maximização de influência em redes complexas para o modelo de spreading activation: Estudo de caso em uma rede de telefonia.