

# MC908 - Projeto de Pesquisa

## Algoritmos de Maximização de Influência em Redes Sociais

Yvens Ian Prado Porto    e    Lucca Miranda Nunes  
RA 184031    e    RA 230554

### 1 Introdução

Este projeto visa explorar e comparar a eficácia de diferentes algoritmos de maximização de influência em redes sociais, isto é, algoritmos que encontram um conjunto de vértices “influenciadores” cujo objetivo é disseminar o máximo possível uma informação. A competição entre algoritmos será avaliada em termos de alcance de influência, utilizando simulações em uma rede de interação. Serão implementados ao menos sete algoritmos distintos, e a competição será realizada dois a dois, utilizando duas abordagens diferentes para a comparação. Além disso, a eficácia de cada algoritmo será analisada isoladamente, sem competição. As possíveis aplicações deste projeto incluem estratégias de marketing, campanhas políticas, campanhas de saúde pública e disseminação de informações em geral.

### 2 Objetivos

Como já foi dito, o principal objetivo do projeto é a comparação entre diferentes algoritmos de maximização de influência. Para isso, temos como objetivos intermediários:

1. Coletar dados da rede social BlueSky para a construção do grafo de interação.
2. Buscar na literatura e implementar ao menos sete algoritmos de maximização de influência (o que inclui as medidas de centralidade vistas em aula).
3. Desenvolver um simulador para executar as competições entre os algoritmos.
4. Realizar simulações para comparar o alcance de influência dos algoritmos.
5. Visualizar as simulações entre os algoritmos por meio de GIFs animados.
6. Determinar uma hierarquia para a eficácia de cada algoritmo em um cenário de competição.
7. Determinar uma hierarquia para a eficácia de cada algoritmo isoladamente.

## 3 Metodologia

### 3.1 Coleta de dados

Pretendemos utilizar a API da rede social BlueSky para coletar dados de interação entre usuários e construir uma rede a partir desses dados, onde os nós representam os usuários. Ainda não está definido o que constituirá uma aresta entre os vértices (pode ser baseado em interações, seguidores, etc.). Precisamos da opinião do orientador para definir isso.

### 3.2 Algoritmos utilizados

Iremos utilizar algoritmos que possuem como entrada uma rede de interação  $G$ , um inteiro  $K$  e um conjunto de vértices  $S$ . A saída esperada é um conjunto  $I$  de  $K$  vértices tal que  $I \cap S = \emptyset$ , sendo  $K$  uma variável que estudaremos. Iremos explorar algoritmos que aplicam diretamente as medidas de centralidade vistas em aula (selecionando os  $K$  vértices com maior medida), algoritmos conhecidos na literatura e algoritmos personalizados por nós.

### 3.3 Simulador e Competições

A simulação consistirá em  $X$  passos (variável a ser estudada), onde a cada passo um vértice é infectado com uma probabilidade proporcional ao número de vizinhos infectados.

Iremos explorar duas abordagens de competição para cada par  $(A, B)$  de algoritmos:

1. Executar os algoritmos com  $G$  sendo a rede original, um  $K$  comum como entrada e  $S = \emptyset$ ; retirar dos conjuntos de saída os vértices que estão na interseção dos dois; comparar o alcance da notícia após a simulação.
2. Executar o algoritmo  $A$  com  $G$  sendo a rede original,  $K = 1$  e  $S = \emptyset$ ; executar  $B$  com  $G$ ,  $K = 1$  e  $S = \{v_1\}$ , sendo  $v_1$  o vértice escolhido por  $A$  na rodada anterior; executar  $A$  com  $G$ ,  $K = 1$  e  $S = \{v_1, v_2\}$ , sendo  $v_2$  o vértice escolhido por  $B$  na rodada anterior e assim por diante, até escolhermos  $K$  vértices com cada algoritmo.

Além disso, iremos analisar também os algoritmos isoladamente, sem competição.

### 3.4 Visualização das Simulações:

Utilizaremos ferramentas de visualização de grafos e GIFs animados para visualizar a propagação da influência durante as competições entre os algoritmos, além de gerar gráficos que mostram o número de vértices infectados em função do número de passos da simulação.

## 4 Avaliação de Resultados

Para avaliar os resultados do projeto, iremos comparar o alcance de influência de cada algoritmo por meio de simulações e análises estatísticas. Utilizaremos métricas como número de vértices infectados e tempo de propagação, além de visualizar os dados através de GIFs animados. A eficácia de cada algoritmo será classificada de duas formas, como um torneio, com competições diretas e em análises isoladas. Assim, poderemos analisar o algoritmo mais eficaz em diferentes contextos e tentar aprimorá-los para terem os melhores resultados possíveis.

## 5 Planejamento e Cronograma

Fase	Atividade	Datas Planejadas
Coleta de Dados	Coletar dados da rede social BlueSky e construir rede de interação	27/09/2024 - 03/10/2024
Implementação	Implementar os algoritmos de maximização de influência	27/09/2024 - 10/10/2024
Desenvolvimento	Desenvolver o simulador para executar as competições	04/10/2024 - 10/10/2024
Execução das Competições	Realizar as simulações e executar as competições entre os algoritmos	11/10/2024 - 24/10/2024
Análise Isolada	Analisar a eficácia de cada algoritmo isoladamente	11/10/2024 - 17/10/2024
Visualização	Criar GIFs animados para visualizar as competições	18/10/2024 - 24/10/2024
Análise dos Resultados	Analisar os resultados das simulações e comparar os algoritmos	25/10/2024 - 31/10/2024
Redação do Relatório	Redigir o relatório final com os resultados e conclusões	04/10/2024 - 06/11/2024