

Análise de comunidades no Instagram

João Antônio Santos Carvalho
Redes Complexas
Ciência da Computação (UFSJ)



1. Introdução

Este trabalho explora a rede de seguidores do Instagram utilizando conceitos de redes complexas para entender as interações e influências entre os usuários. Aplicando uma abordagem de análise de dados, o estudo extrai e modela dados da plataforma para investigar a estrutura da rede. Além disso, o estudo busca entender as preferências esportivas dos estudantes de computação da Universidade Federal de São João del-Rei (UFSJ), cruzando informações de seguidores com perfis de torcidas organizadas do Cruzeiro, Atlético e Athletic.

2. Redes Complexas: Contextualização Teórica

Redes complexas são sistemas formados por um grande número de elementos (nós) interconectados por interações (arestas). Tais redes são onipresentes na natureza e na sociedade, desde redes biológicas até redes sociais e tecnológicas. A análise de redes complexas visa descobrir padrões e regularidades que não são evidentes em uma simples inspeção dos dados brutos.

No caso das redes sociais, como as de seguidores no Instagram, estamos lidando com uma rede onde os nós representam usuários, e as arestas representam relações de "seguir". A compreensão das propriedades desta rede pode fornecer insights valiosos sobre a formação de comunidades, a disseminação de informações e o papel dos influenciadores.

3. Contexto

Este trabalho se insere no contexto da análise de redes sociais, mais especificamente na análise de redes de seguidores do Instagram. Em um cenário onde as redes sociais desempenham um papel fundamental na comunicação e na disseminação de informações, a análise das interações entre usuários pode revelar insights importantes sobre a formação de comunidades, a influência de determinados usuários, e o comportamento coletivo. Por exemplo, em campanhas de marketing digital, entender a estrutura de uma rede de seguidores pode ajudar a identificar influenciadores

chave, ou em estudos sociológicos, pode revelar como grupos se formam e interagem online.

2. Objetivo

O objetivo deste trabalho é realizar um estudo empírico de um contexto real usando ferramentas de análise de redes. Especificamente, o estudo visa analisar a rede de seguidores do Instagram de um conjunto específico de usuários. A partir dessa análise, pretendemos explorar características estruturais da rede, identificar comunidades, e calcular métricas de centralidade para entender melhor a dinâmica de interações entre os usuários.

Adicionalmente, o estudo busca identificar quais times de futebol são mais apoiados pelos estudantes de computação da UFSJ. Para isso, serão coletados os seguidores do perfil do usuário @cacompufsj, que serve como intermediário para identificar esses estudantes. Posteriormente, será investigado quais desses seguidores também seguem as torcidas organizadas de alguns times mineiros: Cruzeiro, Atlético Mineiro e Athletic Club.

3. Metodologia: Como foi feito o trabalho?

Parte 1: Descrição dos Dados + Modelagem

Extração dos dados

Os dados foram obtidos diretamente do Instagram, por meio de raspagem de dados (web scraping), coletando informações sobre os seguidores de usuários específicos da plataforma.

O código coleta os identificadores de perfil dos seguidores de cada usuário, constituindo a base para a modelagem da rede. Informações pessoais dos seguidores, como nomes, fotos de perfil, ou qualquer outro dado que não o identificador único de usuário, não foram coletadas para este estudo.

A extração foi realizada utilizando a biblioteca Selenium para automação de navegação web, permitindo a interação automática com a interface do Instagram. Scripts Python foram desenvolvidos para automatizar o processo de login e navegação nos perfis.

Devido às restrições e políticas de privacidade do Instagram, foi necessário considerar um tempo de espera adequado entre as requisições para evitar bloqueios automáticos da plataforma. Além disso, os dados foram armazenados em arquivos de texto simples (TXT) e a raspagem foi realizada em um ambiente de emulação mobile para imitar o comportamento de um usuário real.

Portanto, por causa das limitações e restrições, para essa pesquisa coletamos apenas 300 seguidores de cada conta. Para evitar muita aleatoriedade, o código faz uma busca entre os seguidores conhecidos primeiro, nesse caso, os estudantes de computação.

Modelagem da Rede

A rede foi modelada como um grafo não direcionado, onde cada nó representa um usuário (tanto os usuários principais quanto seus seguidores), e cada aresta representa uma relação de seguimento entre dois usuários.

Os vértices são os identificadores dos usuários do Instagram que fazem parte da rede. As arestas representam a relação de seguimento entre os usuários, conectando cada usuário ao seu seguidor.

Parte 2: Experimentos

Análise descritiva básica

- **Nós:**

Os nós representam os usuários do Instagram que seguem as contas relacionadas às comunidades identificadas (CACOMP, Athletic, Cruzeiro, Atlético Mineiro). Cada nó no grafo é um seguidor.

- **Arestas:**

As arestas indicam uma relação de seguimento entre os nós, ou seja, a ligação entre dois nós mostra que um usuário segue mais de uma das contas analisadas. A presença de arestas entre nós que pertencem a diferentes comunidades sugere sobreposição de interesses.

- **Densidade:**

A densidade do grafo refere-se à proporção de arestas existentes em relação ao número máximo possível de arestas. Considerando a visualização, o grafo

parece ter uma densidade baixa, sugerindo que a maioria dos usuários seguem apenas uma conta, o que é evidenciado pelas grandes concentrações isoladas de nós.

- **Distribuição de Graus:**

A distribuição de graus no grafo parece ser assimétrica, com a maioria dos nós tendo um grau baixo (seguem apenas uma conta), enquanto um pequeno número de nós tem graus mais elevados (seguem múltiplas contas). Este tipo de distribuição é comum em redes sociais, onde poucos usuários seguem várias comunidades ao mesmo tempo.

- **Coeficiente de Clustering:**

O coeficiente de clustering mede a tendência dos nós a formar triângulos (conexões fechadas) no grafo. Neste caso, os grupos maiores parecem ter coeficientes de clustering baixos, pois a maioria das conexões são entre um usuário e uma única conta, com poucas interconexões entre diferentes comunidades.

Análise Topológica do Grafo

- **Comunidades:**

O grafo mostra quatro grandes comunidades principais (CACOMP, Athletic, Cruzeiro, Atlético Mineiro), identificadas pelas cores no grafo. As comunidades são bem definidas, com pouca sobreposição entre elas. Além disso, há pequenas subcomunidades que conectam duas ou mais das principais comunidades, sugerindo usuários com interesses mistos.

- **Disseminação:**

A disseminação de informação ou influência no grafo provavelmente seria limitada dentro de cada comunidade, mas as subcomunidades (ou pontes) têm o potencial de atuar como canais de disseminação entre as comunidades principais. A presença dessas subcomunidades indica que há pontos de contato onde a informação pode atravessar as fronteiras entre grupos distintos.

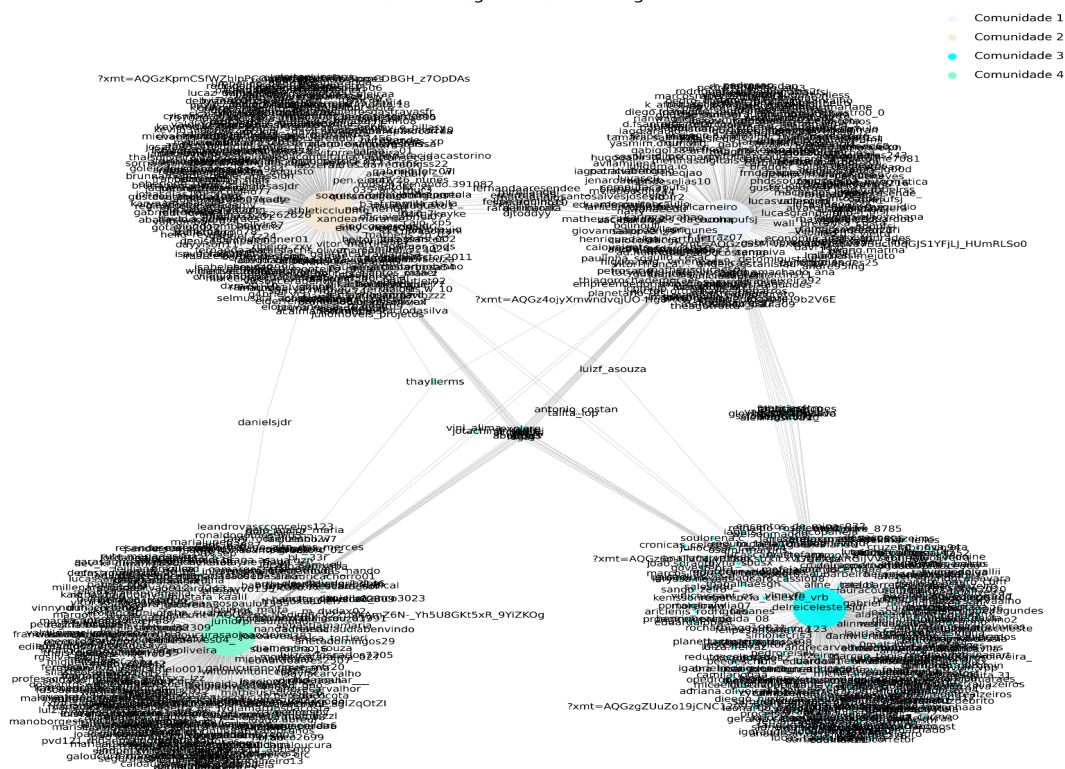
- **Centralidade:**

A centralidade no grafo pode ser observada nos nós que conectam diferentes comunidades. Esses nós servem como "pontes" e têm alta centralidade,

Análise Esportiva:

- Discussão: O que pode ser observado a partir dos resultados?**

Rede de Seguidores do Instagram



Conforme observado na análise visual:

- **Quatro polos principais:** Estes são os usuários que seguem apenas uma das contas analisadas, formando comunidades densas e bem separadas.
- **Três concentrações menores:** Estas são as subcomunidades que conectam duas ou mais das principais comunidades, evidenciando simpatias cruzadas e interesses sobrepostos entre grupos diferentes.

Assumindo que os estudantes da computação estão sendo representados pelos seguidores do CACOMP, podemos observar:

1. A concentração de cima mostra uma quantidade relevante de seguidores em comum das contas do Athletic e do CACOMP, mostrando uma simpatia entre os alunos e o time de São João del Rei
2. O grupo da direita mostra uma bolha de alunos da computação que também seguem a torcida do Cruzeiro de São João del Rei
3. Há uma pequena concentração no meio, onde mistura a comunidade de Cruzeirenses que simpatizam com o Athletic e estudantes da computação que seguem a torcida do Atlético de São João del Rei
4. Além dessas concentrações, há arestas isoladas que representam relações incomuns, como seguir todas as contas estudadas, por exemplo.

4. Conclusões

Este estudo conseguiu extrair e analisar dados dos seguidores do perfil @cacompufsj, modelando a rede de interações e realizando diversas análises para entender as preferências esportivas dos estudantes de computação da UFSJ. O estudo também identificou a relação entre os torcedores do Athletic Club e sua possível simpatia por outros times grandes da região.

Não foi possível realizar a análise de redes muito grandes devido a limitações de tempo e recursos computacionais, além de restrições impostas pelo Instagram para a coleta de dados em larga escala.

No futuro, poderia ser interessante expandir essa análise para incluir dados de outras universidades ou cursos, ou até mesmo comparar as preferências esportivas em diferentes períodos do ano, observando como eventos esportivos influenciam essas redes.re