

Influenciadores no Twitter

Filipe José Maciel Ramalho¹[114.032.785]

Universidade Federal do Rio de Janeiro, Rio de Janeiro RJ, BR
<http://ufrj.br>

Abstract. O mundo vive em meio a uma quantidade de informações impensável a alguns anos atrás. Esta facilidade, não só de acesso a informação, mas também de expressão do pensamento, contribuiu e (ainda contribui) bastante para o desenvolvimento da sociedade como vemos hoje. Agindo como uma das redes sociais mais populares dos últimos tempos, com informações que chegam com uma facilidade até assustadora às pessoas através de suas mensagens curtas, o Twitter é uma plataforma que se mostra favorável à diversos estudos, principalmente os relacionados a sociedade. O objetivo deste trabalho é implementar um método de análise que seja capaz de classificar os usuários, de forma a identificar quais usuários são influenciadores ou não dentro do Twitter. Como não é possível analisar toda a rede, por causa do seu tamanho, a intenção é coletar uma quantidade de dados e relações que represente bem o emaranhado que essa rede representa.

Keywords: Twitter · Influencers · Ranking.

1 Introdução

A sociedade possui um comportamento histórico, onde as pessoas se relacionam com aquelas que possuem as mesmas intuições e as mesmas crenças, no sentido mais amplo da palavra. Assim, este comportamento, em específico, mesmo que em tempos diferentes, já que estamos falando da sociedade contemporânea, continua o mesmo. A digitalização revolucionou muito a maneira como a sociedade vive, modificando comportamentos desde que o primeiro computador foi criado. Porém um outro aspecto importante neste contexto foi o sentido inverso desta mudança: o quanto a sociedade e suas interações modificaram a maneira como a computação, em si, funciona.

Como plataformas de troca de mensagens e interação social, as redes sociais desempenham um papel importante de comunicação que nenhum outro meio de comunicação até seu surgimento tinha conseguido alcançar. O fato de redes conseguirem repassar palavras, sons e imagens em fração de segundos, as tornaram muito atrativas, e a adesão a esta nova forma de comunicação foi muito grande no seu surgimento. Um outro ponto que facilitou esta adesão, foi a democratização da internet, fazendo com que o acesso a web fosse cada vez mais difundido.

Uma destas plataformas, o Twitter, é um facilitador de manipulação das informações que nele existem, provendo alguns recursos para trabalhar de forma

programática as informações que circulam sob sua imensa base de dados. Com essa grande rede de dados, fica mais fácil, também, fazer estudos estatísticos e humanísticos que antes simplesmente não havia uma maneira de realizar. Dentro desta rede de usuários, existe, assim como na vida real, uma certa hierarquia de pensamentos e tipos de pessoas, que definem como o fluxo de informação dentro desta rede social acontece. Então, ao realizar estudos dentro de uma rede como o Twitter, estamos de certa forma, realizando um estudo da sociedade como um todo.

2 Trabalhos relacionados

Existe uma gama muito rica de trabalhos que utilizam o Twitter como dataset para validação de conceito. Existe também uma quantidade considerável de trabalhos que tentam classificar em termos de influência os usuários desta rede. Este estudo tem como base o artigo "Structural Aspects of User Roles in Information Cascades" [1], apresentado na conferência WWW de 2017. Este artigo aborda o tema influência em redes sociais de maneira a agrupar pessoas em quatro grupos distintos, utilizando algoritmos de clusterização, tenta classificar as pessoas nas seguintes categorias:

- **Stars:** São aqueles que postam conteúdo original e não possuem tantos amigos quanto seguidores.
- **Information Sources:** Postam conteúdo original, possuem um maior engajamento em conversas e tem tantos amigos quanto seguidores.
- **Daily chatters:** Geralmente falam das suas rotinas diárias ou replicam informações existentes, não são tão populares.
- **Listeners:** São geralmente apenas consumidores de informação, têm tendência a seguir outras pessoas muito mais que serem seguidos.

O artigo utiliza-se de clusterização através de K-Means e E.M. para seus experimentos, com um sucesso maior para o método E.M.

Outro trabalho realizado, o artigo "Measuring Influence on Twitter" [2], fez uma análise estatística em cima de dados já coletados e processados por ferramentas de terceiros (Klout e Twitter Grader). O objetivo deste trabalho foi de realizar demonstrações das ferramentas e um overview de como as interações do Twitter são feitas, tentando identificar comportamentos e classificá-los de acordo com os seguintes aspectos:

- **Compliance:** Sejam dois usuários definidos como A e B, o usuário A está em "compliance" com o usuário B se A publicamente concorda com as opiniões de B. Opiniões discordantes, se houverem, são mantidas para ele.
- **Identification:** Sejam dois usuários definidos como A e B, o usuário A se identifica com o usuário B se A segue B porque B é uma figura de alta reputação na rede.
- **Internalization:** Processo de aceitar crenças ou comportamento públicos e privados

O artigo propõe também o conceito do indicador SNP (Social Network Potential), que é uma métrica de avaliação que utiliza-se da média de duas outras métricas: a taxa de interação (interação direta entre usuários em uma conversa, por exemplo) e a taxa de menção e retweets.

2.1 Influência e seus impactos

Segundo o artigo "Understanding Twitter Influence in the Health Domain: A social-psychological contribution" [3] a influência é definida a partir da extensividade e da intensividade. É importante entender bem o conceito de influência mais a fundo para que possamos de fato identificá-la em qualquer ambiente, principalmente no Twitter.

A primeira definição, quanto a extensividade, é sobre o alcance que determinada mensagem da qual se queira avaliar a influência, tem. No Twitter, esta avaliação pode ser realizada medindo a relação "seguindo" por "seguidores", número de retweets, número de menções, número de replies, etc. Porém é muito difícil dizer quais destas medidas é mais precisa ou se alguma é realmente precisa. Segundo o próprio artigo, mensagens podem ser distorcidas conforme são repassadas ao longo da rede. Um retweet pode ser feito, por exemplo, ocultando uma questão vinculada ao tweet original, modificando completamente seu sentido, podendo também mudar como se dá sua influência.

Um problema que pode ser observado em relação a uma outra métrica de extensividade é a taxa de seguidores e seguindo, uma medida também utilizada no segundo artigo apresentado neste texto, mas que também possui uma falha bem clara: esta taxa não representa necessariamente se uma pessoa é influente ou não, podendo ela ser alta, mas as interações reais entre o usuário sob estudo e outros usuários, serem bem baixas.

Enquanto a primeira definição é sobre quantas pessoas uma mensagem pode alcançar, a segunda é o quanto a mensagem é relevante para aqueles que a lêem. Esta definição é muito mais subjetiva que a primeira e, por consequência, muito mais difícil de coletar indicadores. Dentre as medidas relacionadas a intensividade, estão a análise de sentimento, Temas e conteúdo ao qual a mensagem está relacionada, e o conceito de retórica - um conceito que nos diz o quanto uma determinada mensagem foi idealizada especificamente para convencer, como por exemplo propagandas.

Cada mensagem é única do ponto de vista do que se quer transmitir e o efeito que realmente causa nas pessoas. No caso do artigo citado, o objeto de estudo são como as mensagens dentro do contexto da pandemia de H1N1 impactavam as pessoas. Um estudo como este é especialmente útil para identificação de fake news, e no contexto da pandemia do Sars-CoV-2, avaliações das redes e sobre como o fluxo de informações impactam as pessoas, podem se mostrar efetivas para prevenir tais ataques digitais e, neste caso, contribuir para superação da doença.

2.2 Papéis na rede

Como já citado neste texto, o artigo que serviu como referência para este trabalho, precisou estipular categorias para classificar as pessoas. Um outro artigo também tenta classificar os usuários do Twitter em grupos de influência: "Identifying Communicator Roles in Twitter" [4].

Enquanto o artigo base tenta partir dos dados a fim de classificar e agrupar os usuários em diferentes categorias, já mencionadas, o presente artigo, tenta, a partir da classificação, encontrar os grupos de usuários. Porém os tipos de classificações são bem semelhantes (apesar de não inteiramente iguais).

- **Idea Starter:** Bem parecido com a definição "Stars", são, em geral, quem iniciam um tópico de discussão, e também os mais engajados em mídias diferentes.
- **Amplifier:** Tem uma rede de conexões muito complexa e em geral compartilham opiniões alheias. "Idea Starters" se tornam famosos, em muitos casos por conta dos "Amplifiers". Aplicadores não tendem a sintetizar informações, eles simplesmente compartilham-nas.
- **Curator:** São questionadores. Tendem, através das ideias de "Amplifiers" e "Idea Starters", contruir novas ideias a partir de questionamentos e novas discussões.
- **Commentator:** Assim como "Curators" e "Amplifiers", "Commentators" tendem a reutilizar ideias de outros. Todavia, tendem a refiná-las, deixando-nas mais claras ao entendimento, sem modificar o conceito inicial.
- **Viewer:** São, assim como a definição de "Listeners", consumidores passivos de informação.

A classificação acima é construída a partir do trabalho de Jonny Bentwood Edelman, um dos autores do artigo, sobre influência no contexto social. O texto também menciona que com este estudo, Edelman produziu o "Tweet Level", uma ferramenta que a partir das classificações mencionadas retorna um número representativo da influência, popularidade e engajamento de um tweet.

2.3 Page-Rank como classificador de usuário

Até aqui foi demonstrado que existem muitos métodos para realizar a resolução do problema da classificação de usuários quanto a influência no Twitter. Porém, uma das maneiras mais simples e ainda assim eficientes de realizar esta tarefa, é utilizando o Page-Rank. O método do Page-Rank para medir influência é o método utilizado como base pelo trabalho que originou o artigo "Social Network Influence Ranking via Embedding Network Interactions for User Recommendation" [5].

Com base neste algoritmo, os autores do artigo propuseram uma nova formulação do próprio Page-Rank especificamente feita para classificação de usuários no Twitter, e a esta nova versão, deram o nome de EIRank. A diferença do EIRank para o Page-Rank original consiste no fato de que o primeiro considera a proximidade entre usuários e utiliza este valor associado ao ranqueamento para determinar influência.

3 O trabalho

Como o Twitter é uma plataforma muito rica em dados sobre como as interações entre as pessoas se dão, a grande maioria dos trabalhos que necessitam de informações desse tipo, são feitos coletando informações desta plataforma. O que pretende-se fazer com este trabalho não é diferente, porém não se perde generalidade, uma vez que em outras redes, reais ou virtuais, um comportamento muito semelhante das pessoas pode ser observado.

A proposta que será apresentada é, na verdade, uma validação deste conceito vastamente abordado por inúmeros outros trabalhos (inclusive os já apresentados durante este texto): a medida de influência. Influenciar é muito útil para os seres humanos, pois somos comunitários, e conseguir convencer e influenciar pessoas pode significar alcançar posição de prestígio. A idéia proposta então é conseguir coletar dados do Twitter, pré processá-los e tentar rankear os usuários com maior capacidade de influência nesta rede, tentando também identificar a área ou tema em que tais usuários são influentes.

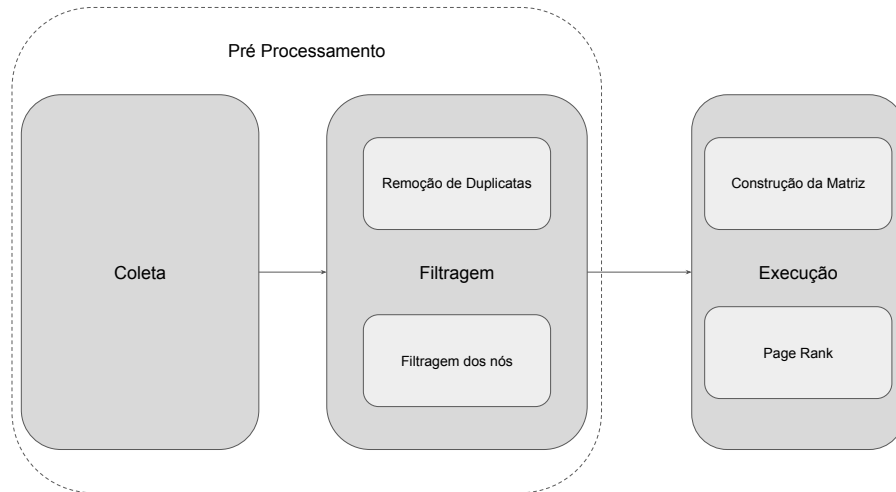


Fig. 1. Fluxo de execução das etapas do experimento

3.1 Coleta de dados

O projeto de rankeamento iniciou-se com a escolha de dataset, como os dados do Twitter seriam coletados e como seriam processados a partir daí.

A plataforma Twitter é bastante razoável em termos de disponibilidade de dados para estudos e implementações diversas, porém, obviamente, ela não pode ser irresponsável com os dados que possui. Então, a plataforma não permite

que desenvolvedores compartilhem seus dados pela internet de forma leviana, limitando-os através de seus termos legais de uso, sob pena de terem o acesso negado aos dados e sofrerem também punições legais. Assim, não deixando que o compartilhamento de dados de seus usuários saiam do domínio da plataforma de maneira desordenada.

Como o compartilhamento sendo bastante restrito, não foi possível encontrar datasets interessantes na internet contendo tweets bem interconectados que contivessem também os usuários que os criaram e o conteúdo dos textos, que eram requisitos para a execução do projeto. Portanto, a solução foi tentar criar o próprio dataset a partir de consultas a api oficial da plataforma[6].

O desafio enfrentado a partir da decisão de coletar dados diretamente da api, foi a questão de limitação de requests que a própria api possui, não permitindo que usuários realizem um número alto de requisições em um determinado intervalo de tempo (em geral 15 minutos). Isso era inviável para o trabalho, que necessitava coletar o máximo de dados em um intervalo de tempo curto por causa do tempo limite de entrega de resultado. Até que com a ajuda da ex-aluna de graduação da UFRJ, Thaís Luca, uma das rotas descritas na documentação da api pareceu bastante promissora para o objetivo do experimento, a rota HTTP denominada "Sampled stream".

A rota "Sampled stream" é um endpoint que possui um limite muito pouco restrito, e é especialmente feita para alimentar, com dados em massa, aplicações. Esta rota fornece dados de 1 por cento de todo o conteúdo mundial do Twitter de forma aleatória. O tamanho final do arquivo gerado foi de 391,8MB (647645 tweets). O código da coleta foi construído em NodeJS, que possui um bom suporte para realizar requisições HTTP e sem muitos problemas para desenvolvê-lo.

3.2 Refinamento dos dados

Depois de escolhida e realizada a forma de coleta de dados, o número de tweets coletados era muito grande e a maioria deles não continham nenhuma relação com outros tweets presentes na coleção. Assim, foi necessário um pré-processamento para a retirada destes tweets isolados na coleção.

Com isso, foi construído um script para o refinamento desses dados. Construído também utilizando NodeJS, o script de refinamento tem a tarefa de ler o arquivo gerado na etapa de coleta, remover eventuais tweets duplicados e separar aqueles tweets que fazem referência a outros tweets dentro da coleção, ou que sejam referenciados por outros tweets também dentro da coleção. Os tweets que tem alguma referência, porém não dentro da coleção, infelizmente não são razoáveis para o processamento do experimento no contexto do presente trabalho, pois seria necessário a busca do tweet referenciado na api do Twitter, o que é impraticável devido às suas limitações para a versão standard (que foi a versão utilizada).

O script de refinamento, então, foi construído para separar o conteúdo do dataset inicial em dois novos arquivos: um deles contendo todos os tweets com alguma relação dentro da própria coleção, e outro contendo todos os demais.

3.3 Execução do Page-Rank

Após o refinamento dos dados, o tamanho do arquivo de dataset gerado passou a ter 4202 tweets, totalizando 2.4MB de espaço de armazenamento. A etapa que precisa consumir estes dados precisa, portanto estar escrita em uma linguagem que possua bom desempenho de processamento.

O algoritmo do Page-Rank utilizado para o experimento é o mesmo apresentado em sala de aula, utilizando o artifício de teleporte. No entanto, tal implementação não se preocupa com os becos sem saída do algoritmo, pois tais becos são removidos na etapa de pré-processamento. De maneira geral, o mapeamento dos nós do Page-Rank com relação a coleção é definida de maneira que os usuários, a quem cada tweet pertence, são os nós do grafo. As arestas do grafo são definidas como sendo cada retweet, reply ou qualquer tweet que tenha referência clara a um outro tweet, de um outro determinado usuário, ou do mesmo usuário (neste último caso, mapeando para uma aresta que sai e chega no mesmo nó). A fórmula do Page-Rank utilizado no projeto é apresentada na equação (1) abaixo. O parâmetro beta utilizado foi de 80% e a diferença de resultados entre iterações para considerar convergência foi de 0,1 unidades entre somatórios dos elementos dos vetores gerados.

$$PR(a) = \frac{(1 - \beta)}{T} + \beta \sum_{i=1}^n \frac{PR(p_i)}{L(p_i)} \quad (1)$$

A construção da matriz de incidências precisava então extrair dos tweets, todos os usuários a que eles pertenciam, para criar a lista de usuários a serem testados. As incidências precisavam ser coletadas avaliando cada tweet e com qual ou quais outros tweets ele se relacionava. O algoritmo então extrairia os usuários tanto dos tweets referenciados quanto do tweet que estivesse fazendo a referência, adicionando toda essa informação de forma sintetizada, como incidências entre nós, na matriz.

Inicialmente o algoritmo para execução do Page-Rank foi escrito utilizando a linguagem C++, pois não estou habituado a trabalhar com a linguagem Python. Entretanto, depois de muitas tentativas, sem sucesso, para conseguir lidar bem com o tamanho do conjunto de dados, que por si só já era construído em uma estrutura complicada para o C++ (uma lista de JSONs), resolvi utilizar a linguagem Python para realizar a tarefa.

Após a execução do algoritmo do Page-Rank, um vetor de valores seria produzido, de forma que tal vetor é um autovetor da matriz criada a partir dos critérios do Page-Rank. Este autovetor na verdade indica quais usuários da coleção que têm seus tweets mais referenciados. Então, após coletar esta lista de usuários, foi executada uma ordenação dos ids de usuários da coleção para que houvesse uma lista ordenada do mais influente para o menos influente dentro do conjunto de teste.

4 Resultados

Após executar todas as etapas do experimento é possível perceber quem são os mais influentes e quem são os menos influentes dentro da coleção. O tempo que o algoritmo de ranqueamento levou para processar todas as operações até convergir foi em torno de 1.18 segundos. A construção da matriz durou em torno de 0.05 segundos, enquanto o tempo de carregamento do arquivo de dados foi de aproximadamente 0.04 segundos.

```
{
  "data": {
    "id": "1322238703843069959",
    "lang": "tr",
    "author_id": "1248136372013105152",
    "referenced_tweets": [
      {
        "type": "replied_to",
        "id": "1322233712872984577"
      }
    ],
    "text": "@Doqannrrr Cok gecmiş olsun 🙏rabbim birdaha yasatmasın🙏",
    "includes": {
      "users": [
        {
          "id": "1248136372013105152",
          "name": "Hafza Naz",
          "username": "hafza_naz"
        },
        {
          "id": "1317120794972065793",
          "name": "Aşkî Tevekkül (AişeNur)",
          "username": "Doqannrrr"
        }
      ]
    },
    "tweets": [
      {
        "id": "1322233712872984577",
        "lang": "tr",
        "author_id": "1317120794972065793",
        "text": "Mekan aynıı, 🙏sabahlıcazz https://t.co/H15RBnrum4"
      }
    ]
  }
}
```

Fig. 2. Exemplo de tweet no formato JSON retirado de uma linha da coleção

A figura 2 apresenta o que seria uma linha dentro do arquivo de dados adquirido através da api do Twitter. Ela nos mostra um JSON contendo dados sobre o tweet escolhido como exemplo, de id "1322238703843069959". Processar milhares de JSONs como este não é uma tarefa simples, e isto foi evidenciado quando houve uma tentativa de executar o algoritmo em python com a coleção completa de 647645 tweets apenas para teste de eficiência do algoritmo, sem se preocupar com o resultado. Quando o algoritmo tentou construir a matriz, utilizando este conjunto grande, ocorreu uma exceção onde o terminal informava que tentou-se sem sucesso alocar 2TB em memória. O motivo da tentativa de alocar 2TB de memória, ao que se acredita, é por causa do fato de a complexidade de espaço do algoritmo ser de $O(n^2)$.

As tabelas 1 e 2 mostram parte do resultado final do ranking de usuários contendo também o número de tweets de cada um dos usuários listados bem como o número de referências feitas a cada um dos usuários, entendendo como referência: retweets, menções, replies, etc. Então por exemplo, o usuário *itsJeffTiedrich* fez 1 tweet (que pode ser também um retweet) apenas mas foi referenciado 9 vezes por outros tweets de outros usuários (ou dele mesmo). A tabela 1 lista os 20 primeiros usuários ranqueados no Page-Rank enquanto a tabela 2 mostra os 20 últimos.

Table 1. Dados dos 20 rankeados como mais influentes, na ordem em que aparecem (O primeiro é o mais influente).

Nome de usuário	Nº de tweets	Nº de refs.	Nome de usuário	Nº de tweets	Nº de refs.
ryo_maru4	1	1	StormUsmc	1	1
bitsofjupiter	1	1	8g_k_	1	1
JacksonWGlobal	1	2	oldandloaded	1	1
thipchaanok	1	1	itsJeffTiedrich	1	9
nctsupportint	1	3	archivegoat	1	2
k2rll	1	1	filtrmygxz	1	1
sa_jxr	1	1	memeghello	1	1
watchmeloko	1	1	MusicBNK48THFan	1	2
arreaaortega	1	1	commentary_a	1	1
tamborzinho_	1	1	outshinexo	1	1

Table 2. Dados dos 20 rankeados como menos influentes, na ordem reversa em que aparecem (O último é o menos influente).

Nome de usuário	Nº de tweets	Nº de refs.	Nome de usuário	Nº de tweets	Nº de refs.
Claudio_Mir	2	1	yyexxr2	1	0
konyasukon	2	1	GumGumRequiem	2	1
s_ben212	2	2	Andy7_hdz	2	1
AKMU_STAR	2	1	interrodang	2	1
09isyourangel	2	3	thomasthetylar	2	1
crissyyanggg	2	1	iamHaqem	2	1
JZbzzz	2	1	dongjunk2m	2	1
btrzamomattel	2	1	prin_mariella	2	1
929ftwalls	2	1	NorthEastDundee	2	1
rutosthetic	2	2	Alpim1501	2	1

5 Conclusão e Perspectivas futuras

Pôde-se verificar com o estudo previamente realizado neste trabalho, que a análise de influência é extremamente necessária para o estudo da sociedade como um todo e que este tipo de estudo revela não somente quem é ou não influente em uma rede, mas de maneira mais bem elaborada, pode ajudar a identificar pessoas que possam ser prejudiciais a outras bem como o aproveitamento desta informação para veicular informações importantes para que impactem o maior número possível de pessoas dentro da rede destes influenciadores.

Com a pobreza em datasets públicos desta natureza, foi necessário construir um dataset exclusivamente para o experimento, porém tal dataset se mostrou insuficiente para alcançar todos os objetivos pretendidos, o endpoint "Sampled Stream" da api do Twitter não foi uma alternativa satisfatória, apesar de ter sido a única conseguida dentro do tempo de duração do projeto.

Com o dataset construído foi possível fazer uma matriz para a execução do Page-Rank, entretanto tal matriz revelou-se muito esparsa, mostrando que as conexões entre os usuários da coleção é bastante esparsa. Ao visualizar a matriz (figura 3), algo ainda mais surpreendente foi descoberto, a matriz é praticamente diagonal, a menos de algumas células fora do padrão. Este comportamento, no grafo, indica que o formato dele é quase de um grafo linha, sem muitos ciclos. Tomando a api do Twitter como perspectiva, podemos interpretar que este resultado pode significar que, ao contrário do que se supunha sobre esta rota da api, no início do trabalho, ela não fornece dados de forma randômica, ao contrário, ela parece fornecer os dados de forma sequencial no tempo.

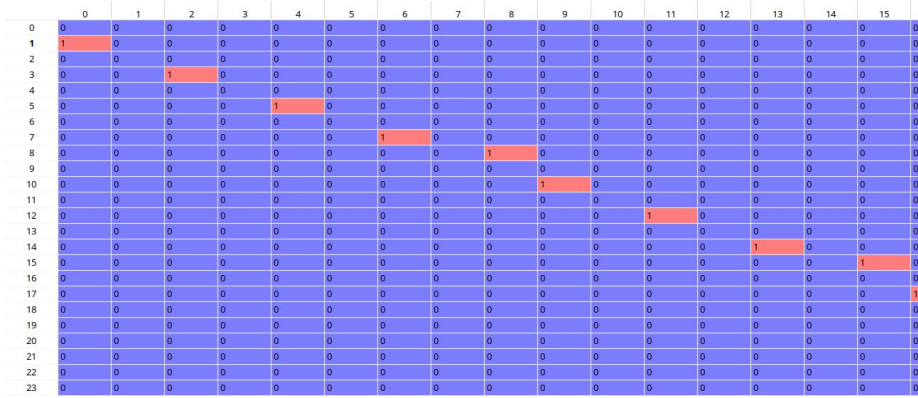


Fig. 3. Visualização de parte da matriz de incidência gerada

O que se pode concluir finalmente sobre a lista gerada com o Page-Rank é que de fato as teorias e suposições mencionadas nos artigos citados neste trabalho se mostram verdadeiras. O resultado da execução do Page-Rank sugere que os usuários que mais bem rankeados, ou seja, os mais influentes, tendem

a postar menos tweets, enquanto os classificados como menos influentes, criam muitos tweets, todavia, como não foi diferenciado se tais tweets são na verdade referências a outros tweets, a intuição é de que criar é mais difícil que replicar, ou seja, há uma probabilidade maior que os tweets dos influenciadores sejam conteúdo original, enquanto os tweets dos influenciados sejam na verdade retweets, replies, etc.

Para futuros trabalhos é, então almejado que se tente reproduzir o experimento novamente, porém buscando alternativas melhores de datasets, acredito que com um tempo maior e mais calma, eventualmente um dataset que se enquadre melhor ao experimento será encontrado.

Outra implementação que não foi possível neste trabalho por causa do baixo volume de relações significativas e também por causa do tempo de duração do experimento, foi a classificação dos influenciadores por área de atuação ou tema de interesse. A ideia parece bastante promissora e com certeza seria importante de ser implementada no futuro.

References

1. Taxidou, Io; Fisher, Peter: Structural Aspects of User Roles in Information Cascade. Em: Proceedings of the 26th International Conference on World Wide Web Companion, 2017, pp. 1505–1509. <https://doi.org/10.1145/3041021.3053906>
2. Anger, Isabel; Kittl, Christian: Measuring Influence on Twitter. Em: i-KNOW '11: Proceedings of the 11th International Conference on Knowledge Management and Knowledge Technologies, 2011 pp. 1–4. <https://doi.org/10.1145/2024288.2024326>
3. McNeill, Andrew R; Briggs, Pam: Understanding Twitter Influence in the Health Domain: A social-psychological contribution. Em: WWW '14 Companion: Proceedings of the 23rd International Conference on World Wide Web, 2014, pp. 673–678. <https://doi.org/10.1145/2567948.2579280>
4. Tinati, Ramine; Carr, Leslie A.; Hall, Wendy; Bentwood, Jonny: Identifying Communicator Roles in Twitter. Em: WWW '12 Companion: Proceedings of the 21st International Conference on World Wide Web, 2012, pp. 1161–1168. <https://doi.org/10.1145/2187980.2188256>
5. Bo, Hongbo; McConville, Ryan; Hong, Jun; Liu, Weiru: Social Network Influence Ranking via Embedding Network Interactions for User Recommendation. Em: WWW '20: Companion Proceedings of the Web Conference, 2020, pp. 379–384. <https://doi.org/10.1145/3366424.3383299>
6. <https://developer.twitter.com/en/docs/twitter-api>, <https://developer.twitter.com/en/docs/twitter-api>, <https://developer.twitter.com/en/docs/twitter-api>. Último acesso em 8 de Novembro de 2020.