

## Bitcoin OTC trust weighted signed network

Link: [SNAP: Signed network datasets: Bitcoin OTC web of trust network](#)

Rede de avaliações de confiabilidade entre usuários da plataforma Bitcoin-OTC.

Usuários deram notas de confiança de  $-10$  a  $+10$  para outros usuários após trocas.

Cada linha do arquivo representa uma avaliação de confiança.

### Dataset statistics

Nodes	5,881
Edges	35,592
Range of edge weight	-10 to +10
Percentage of positive edges	89%

Formato dos dados: | source | target | rating | time |.

Onde:

- SOURCE: id do nó de origem, ou seja, avaliador.
- TARGET: id do nó de destino, ou seja, avaliado.
- RATING: a classificação da fonte para o alvo, variando de  $-10$  a  $+10$  em etapas de 1.
- TIME: o tempo da classificação (medido em segundos desde a Epoch).

Exemplo:

source	target	rating	time	Explicação
1	2	8	—	Usuário 1 avaliou o 2 com nota 8.
3	2	-7	—	Usuário 3 avaliou o 2 com nota -7.

## Problemas com BFS e DFS em Redes de Confiança (Trust Network)

**BFS:** [Jamali, M. e Ester, M. (2009)] - Using a Trust Network to Improve Top-N Recommendation

<https://www2.cs.sfu.ca/~ester/papers/RecSys-2009-TopNRecommendation.final.pdf>

- Usando uma rede de confiança para melhorar a recomendação Top-N.
- Trabalha com sistemas de recomendação (ex.: filmes, livros...). Usuários avaliam itens.
- O sistema deve prever uma avaliação de um usuário  $x$  em um item não avaliado, baseado em avaliações conhecidas.
- Top-N item recommendation: usuário quer uma lista ranqueada de itens que ele ainda não avaliou mas provavelmente daria valores altos (gostaria muito).
- Collaborative Filtering é a abordagem mais popular para criar sistemas de recomendação que podem prever avaliações para um dado usuário ou item, mas não são boas para recomendação Top-N quando relacionada à usuários que avaliaram apenas um número muito pequeno de itens.
- Solução: recomendação baseada em confiança. Faz recomendações com base nas classificações dos usuários que são direta ou indiretamente confiáveis por você.
- Abordagem Combinada (4.2) do artigo:
  - Quando um usuário  $u$  confia em um usuário  $v$ , não necessariamente significa que eles avaliam os mesmos itens, mas sim que, se eles avaliarem os mesmos itens, provavelmente a avaliação será similar.
  - Dois usuários que já avaliaram alguns itens em comum tendem a avaliar mais itens em comum. Portanto, é mais provável que usuários com padrões de avaliação semelhantes aos dos usuários de origem também avaliem o item.
  - Calcula os  $K$  principais usuários confiáveis na rede e classifica os itens avaliados por esses usuários para calcular os  $N$  principais itens recomendados. Usa a abordagem da filtragem coletiva para calcular outro conjunto de  $N$  itens e produz uma lista combinada.

- Essa abordagem utiliza usuários semelhantes, com maior probabilidade de ter os itens retidos, e também utiliza uma rede de confiança para lidar com usuários de inicialização a frio. Observe que metade dos usuários são usuários de inicialização a frio, e a filtragem colaborativa não consegue encontrar usuários semelhantes para esses usuários.
- **Para calcular os K principais usuários confiáveis, pode se utilizar o Breadth First Search (Busca em Largura): para encontrar os k\_2 usuários mais próximos do usuário de origem u na rede de confiança.**
- **A BFS garante que usuários mais próximos (com menor profundidade) são considerados mais influentes na recomendação.**

**DFS:** Lily Z. (shuxing3), Anjana G. (anjanag3), Trisha C. (trishac2), Anagha S. (anaghas2) - Bitcoin-Trust-Analysis

<https://github.com/LilyZ-03/Bitcoin-Trust-Analysis?tab=readme-ov-file>

- Utiliza o DFS, o Componente Fortemente Conectado de Tarjan e o Dijkstra para analisar os dados de classificação de transações de usuários de Bitcoin OTC em um gráfico.
- O objetivo era utilizar o conjunto de dados de redes assinadas ponderadas pela confiança do Bitcoin OTC do SNAP (Stanford) para identificar relacionamentos existentes entre usuários de Bitcoin e, por fim, determinar a confiabilidade do usuário.
- A busca em profundidade (DFS) no grafo foi realizada para desconectar com sucesso as arestas negativas ou "fracamente relacionadas" usando um limite definido como 2 para incluir apenas vendedores com um nível de confiança relativamente bom (3-10, onde 10 é altamente confiável).
- Os testes DFS verificam se todas as arestas com pesos iguais ou inferiores ao limite (definido como 2) são removidas e se todas as arestas restantes têm um peso acima do limite.
- **Basicamente, a DFS é aplicada após a filtragem dessas arestas com confiança inferior a um limite (peso  $\leq 2$ ), com o objetivo de explorar os**

**subgrafos resultantes e identificar componentes acessíveis a partir de um usuário de origem. A busca permite verificar quais usuários ainda podem ser alcançados por caminhos com confiança aceitável, após a remoção das relações fracas.**

- O Algoritmo de Tarjan é aplicado para encontrar componentes fortemente conectados, representando grupos com confiança mútua elevada.
- O Algoritmo de Dijkstra é utilizado sobre o subgrafo resultante para encontrar os caminhos mais curtos (ou com maior confiança acumulada) até potenciais usuários recomendados.
- Por meio do processo, foi observado que as redes de confiança entre usuários de Bitcoin são mais densamente conectadas do que o esperado. Ao implementar com sucesso os algoritmos desejados, foi possível determinar o vendedor mais confiável para recomendar a um usuário específico.

### **Verificação de métricas básicas para comunidades**

1. Número de nós: 5881

Tamanho médio, mas suficiente.

2. Número de arestas: 35592

Parece baixo para a quantidade de nós.

3. Densidade: 0.001029

Extremamente baixa. Redes de confiança raramente têm densidade alta.

Pode indicar comunidades mais fácil?

4. Grau médio: 12.10

Em média, um usuário confia em 6 outros usuários e é confiado por 6.

Rede esparsa.

5. Número de componentes fortemente conectados: 1144

Rede é fragmentada em várias regiões conectadas.

Comunidades relevantes?

## Modelos de Redes

Cálculo do Diâmetro: O digrafo não é fortemente conectado. Existem nós no grafo que não alcançam todos os outros nós seguindo as arestas direcionadas.

- Solução: Calcular o diâmetro do maior componente fortemente conectado.
- Diâmetro: 11.

Coeficiente de Clusterização:

- *average\_clustering*: calcula a média dos coeficientes de clusterização local para todos os nós da rede. A métrica fornece uma medida única do nível geral de agrupamento na rede.
- Clusterização: 0.1510698055090718.

Probabilidade de existir uma aresta entre nós:

- Probabilidade ( $p_{\text{edge}}$ ): 0.0010292571373048454.

Comparação dos Modelos com as métricas encontradas:

Modelo	L (diâmetro)	C (clusterização)
Original	11	0.15107
Grafo Aleatório	10	0.00103
Small-World	7	0.60958
Preferential Attachment	5	0.01317
PA Direcionado	11	0.00316

Small-World (K)	p	L (diâmetro)	C (clusterização)
2	$p_{\text{edge}}$ (0.0010293)	2908	0.0
4	$p_{\text{edge}}$ (0.0010293)	542	0.49899
6	$p_{\text{edge}}$ (0.0010293)	277	0.59851
12	$p_{\text{edge}}$ (0.0010293)	98	0.67992
18	$p_{\text{edge}}$ (0.0010293)	46	0.70399
24	$p_{\text{edge}}$ (0.0010293)	33	0.71554
30	$p_{\text{edge}}$ (0.0010293)	26	0.72233

4	0.01	113	0.48737
4	0.05	35	0.43517
4	0.1	22	0.37449
4	0.5	12	0.07006
<b>4</b>	<b>0.35</b>	<b>14</b>	<b>0.15651</b>
5	0.35	14	0.15651

\*\*\*