

4 Testes e Resultados

Para o desenvolvimento deste trabalho, as estruturas e tecnologia das bases utilizadas para os testes do algoritmo desenvolvido são proprietárias da Stone Age Tech. Todos os campos da base de dados são automaticamente definidos como chaves. Uma vez construída a base, todos os campos podem ser utilizados como chaves de consulta. Não existem mais limitações sobre as consultas que podem ser feitas, nem a necessidade de se pedir que sejam construídas novas agregações ou cubos de análise dos dados.

Todo atributo (ou variável) pode ser cruzado com qualquer outro, dando assim total liberdade para o usuário. Há a possibilidade de se extrair qualquer tipo de informação dos dados disponíveis.

O gerenciador faz uso de algoritmos de compressão de dados reduzindo a necessidade de espaço em disco em 80%. A base de dados final ocupa, em média, 20% do espaço ocupado pelos acervos originais e, com isso, bases de dados que antes requeriam equipamentos de armazenamento avançados e custosos podem agora ser disponibilizadas em discos rígidos de computadores pessoais comuns. Essa característica da base de dados facilitou o desenvolvimento deste trabalho, que investiga mineração de dados em bases com grandes volumes de informação, tanto em relação ao número de registros quanto à quantidade de variáveis.

Para que aplicações possam acessar sistemas gerenciadores de banco de dados diferentes de maneira transparente, independente da linguagem e da base de dados, foi criado um padrão de acesso a esses sistemas que define interfaces de acesso de maneira que não seja necessária a utilização de métodos específicos de cada base, denominado ODBC - *Open Data Base Connectivity*.

A utilização de um conector (driver) ODBC específico para a base de dados que se deseja utilizar permite que as informações sejam obtidas utilizando-se a linguagem SQL - *Structured Query Language*.

4.1. Descrição do Ambiente de Testes – Protótipo

Para testar a eficiência e a eficácia do algoritmo proposto, foi desenvolvido em Delphi um protótipo de testes que, por permitir a escolha de quaisquer conectores

ODBC, é genérico a qualquer tipo de base e não somente às bases com a tecnologia da Stone Age Tech.

Inicialmente, o conector ODBC desejado é escolhido dentre a lista disponibilizada como mostra a figura a seguir.

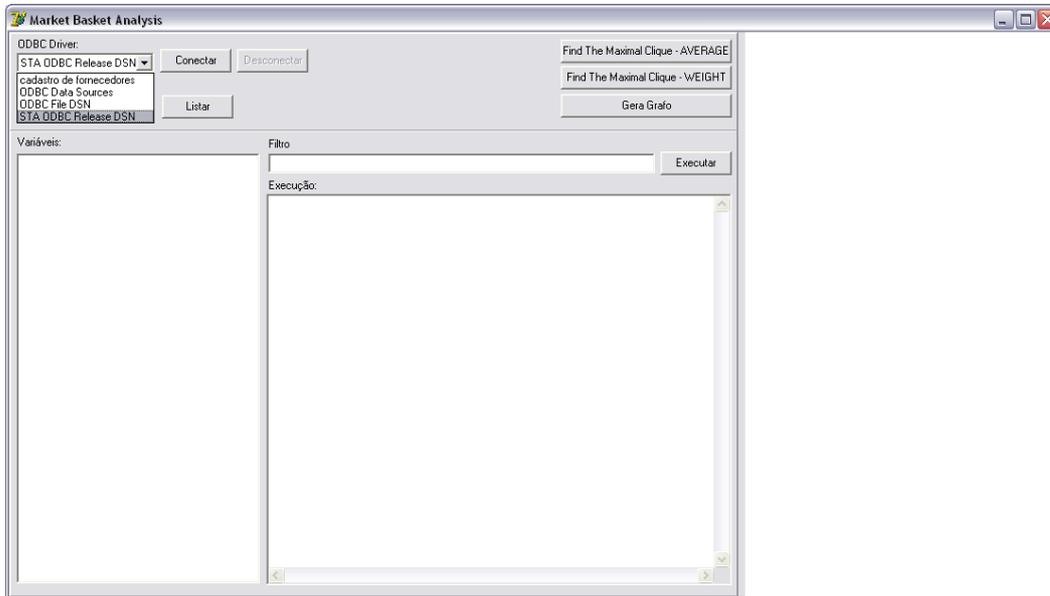


Figura 7 Protótipo: escolha do conector ODBC

Para o acesso aos dados serem disponibilizados, é necessário entrar com nome de usuário e senha válidos na tela de autenticação que será exibida logo após o clique no botão Conectar.

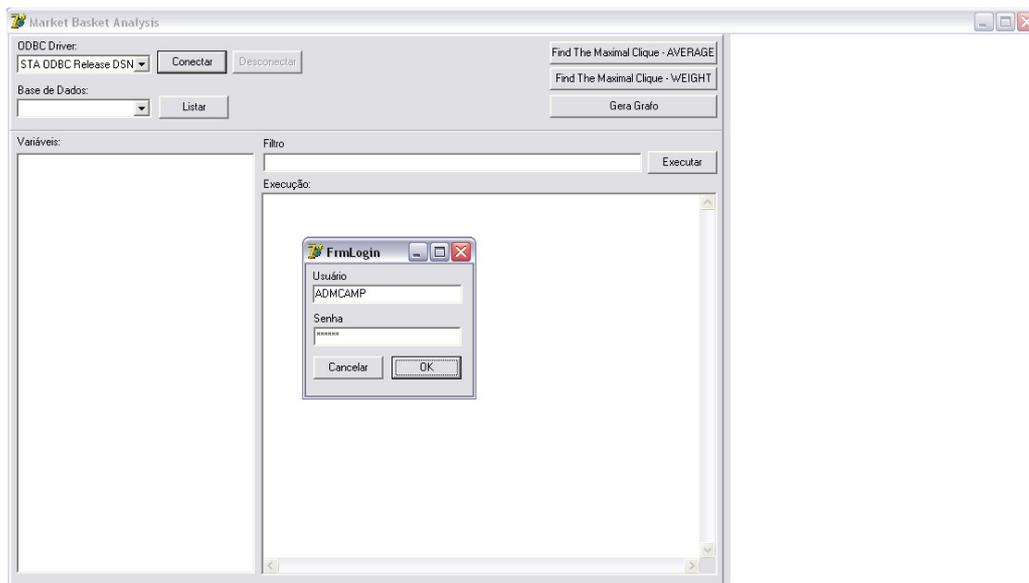


Figura 8 Protótipo: Autenticação no ODBC

Uma vez conectado, a lista de bases será exibida para o usuário, que deve escolher com qual deseja trabalhar. Em seguida, o usuário deve solicitar a listagem das variáveis disponíveis para cruzamento. Neste protótipo estão sendo exibidas todas as variáveis da base. Entretanto, para os estudos contemplados neste trabalho apenas as variáveis discretas serão utilizadas.

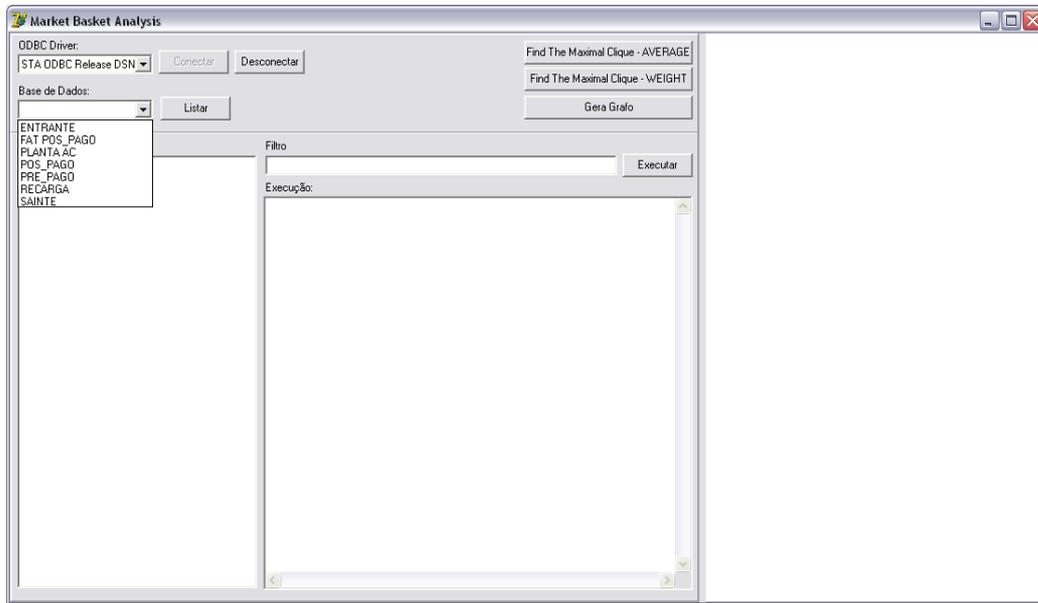


Figura 9 Protótipo: Escolha da base de dados

O usuário escolhe as variáveis sobre as quais deseja estudar as relações de associação e solicita a execução do comando. Se desejar realizar algum filtro nos registros da base em questão, os critérios precisam ser digitados no campo filtro antes do comando de execução, como mostra a figura a seguir.

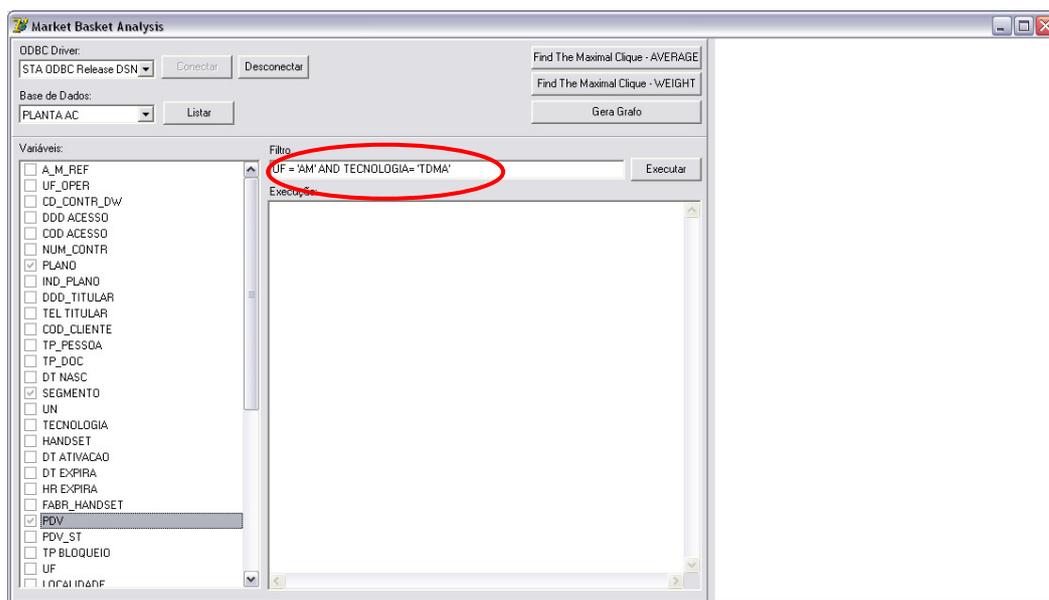


Figura 10 Protótipo: Inclusão de filtro na consulta

Durante o processamento, informações como a consulta (*query*) definida, tempo de execução pelo ODBC em milissegundos, total de registros considerados na consulta e o tempo para montar a matriz de adjacências representando as relações encontradas são exibidas.

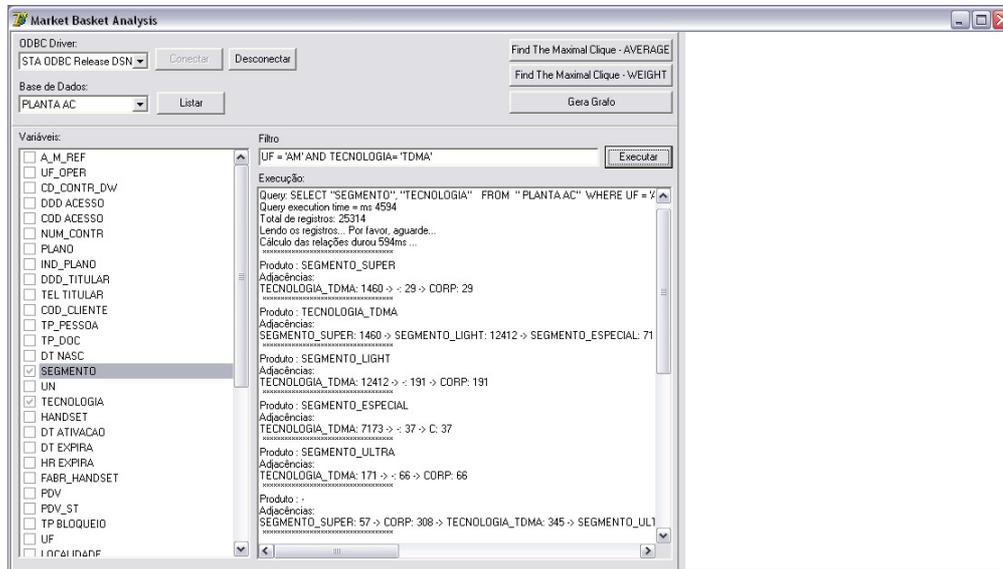


Figura 11 Protótipo: Exibição das informações relevantes

4.2. Testes

Os testes foram realizados sobre uma base, denominada Vendas, que possui 25.119.414 registros e 197 variáveis. Apenas 81 dessas variáveis foram utilizadas nos testes, pois apenas estas são discretas, ou seja, com domínio de valores pré-definidos. Esta base contém as informações de abordagens de marketing para pessoas que já são clientes e para pessoas que ainda não são clientes de uma instituição financeira, como: os produtos a que são elegíveis, as restrições de elegibilidade a campanhas (de acordo com o CPF dos clientes), algumas informações cadastrais, dentre outros.

Para o primeiro caso de teste, foram selecionadas:

- 2 variáveis que representam produtos da empresa proprietária da base de dados
- 2 variáveis que indicam se é a primeira vez que aquele CPF foi indicado para adquirir esses produtos, respectivamente
- 1 variável que indica se o CPF é de uma pessoa que já é cliente da instituição ou que ainda não é, e

- 1 variável com a quantidade de vezes que esse CPF foi abordado em uma campanha de marketing.

Analisando os padrões encontrados pelo algoritmo (ver Quadro 1), verifica-se que as relações mais recorrentes estão entre os registros que representam pessoas que

- não são clientes ainda
- foram abordados mais de 4 vezes por uma campanha de marketing
- são elegíveis a campanhas por mala-direta e telemarketing e
- normalmente não é a primeira vez que se tornaram elegíveis aos produtos selecionados para a análise.

Com isso, pode-se concluir que os indivíduos desta base já foram inúmeras vezes abordados para adquirirem os produtos CITIMAX e CITICARD e, mesmo assim, não aderiram à campanha. Tendo essas informações, os analistas perceberam que precisam repensar o modelo de abordagens desses possíveis clientes e/ou a eficiência das campanhas que estão sendo realizadas, buscando tem mais sucesso com menos abordagens.

```

Query: SELECT "CITIFINANCIA", "FX AB", "CITI MAX", "CITICARD", "CITICARD N", "CITIMAX
N" FROM "CREDICARDCITI_CITI_VENDAS" WHERE SEXO = 'F' OR SEXO = 'M'
Query execution time = ms 74515
Total de registros: 318383
Lendo os registros... Por favor, aguarde...
Cálculo das relações durou 57985ms ...
*****
Produto : CITIFINANCIA_NAO
Adjacências:
FX_AB_NAO: 286581 -> CITI_MAX_NAO: 173888 -> CITICARD_NAO: 245837 ->
CITICARD_N_NAO: 286649 -> CITIMAX_N_NAO: 250698 -> CITI_MAX_ELEG_MD_TLMK: 112761 -
> CITICARD_ELEG_MD_TLMK: 40812 -> CITIMAX_N_PRIMEIRA: 35951 -> FX_AB_4_OU MAIS:
65 -> FX_AB_3: 1 -> FX_AB_2: 1 -> FX_AB_1: 1
*****
Produto : FX_AB_NAO
Adjacências:
CITIFINANCIA_NAO: 1243649 -> CITI_MAX_NAO: 191441 -> CITICARD_NAO: 270110 ->
CITICARD_N_NAO: 318315 -> CITIMAX_N_NAO: 278179 -> CITI_MAX_ELEG_MD_TLMK: 126874 -
> CITICARD_ELEG_MD_TLMK: 48205 -> CITIMAX_N_PRIMEIRA: 40136 -> CITIFINANCIA_SIM:
129475
*****
Produto : CITI_MAX_NAO
Adjacências:
CITIFINANCIA_NAO: 1243649 -> FX_AB_NAO: 1058042 -> CITICARD_NAO: 176691 ->
CITICARD_N_NAO: 191472 -> CITIMAX_N_NAO: 191472 -> CITICARD_ELEG_MD_TLMK: 14781 ->
CITIFINANCIA_SIM: 24995 -> FX_AB_4_OU MAIS: 83 -> FX_AB_3: 1 -> FX_AB_2: 1
*****
Produto : CITICARD_NAO
Adjacências:
CITIFINANCIA_NAO: 1243649 -> FX_AB_NAO: 1058042 -> CITI_MAX_NAO: 559633 ->
CITICARD_N_NAO: 270159 -> CITIMAX_N_NAO: 232936 -> CITI_MAX_ELEG_MD_TLMK: 307142 -
> CITIMAX_N_PRIMEIRA: 37223 -> CITIFINANCIA_SIM: 129475 -> FX_AB_4_OU MAIS: 156 ->
FX_AB_2: 1
*****

```

```

Produto : CITICARD_N_NAO
Adjacências:
CITIFINANCIA_NAO: 1243649 -> FX_AB_NAO: 1058042 -> CITI_MAX_NAO: 559633 ->
CITICARD_NAO: 503094 -> CITIMAX_N_NAO: 278237 -> CITI_MAX_ELEG_MD_TLMK: 307142 ->
CITICARD_ELEG_MD_TLMK: 93524 -> CITIMAX_N_PRIMEIRA: 40146 -> CITIFINANCIA_SIM:
129475 -> FX_AB_4_OU MAIS: 156 -> FX_AB_3: 1 -> FX_AB_2: 1 -> FX_AB_1: 1
*****
Produto : CITIMAX_N_NAO
Adjacências:
CITIFINANCIA_NAO: 1243649 -> FX_AB_NAO: 1058042 -> CITI_MAX_NAO: 559633 ->
CITICARD_NAO: 503094 -> CITICARD_N_NAO: 278237 -> CITI_MAX_ELEG_MD_TLMK: 307142 ->
CITICARD_ELEG_MD_TLMK: 93524 -> CITIFINANCIA_SIM: 129475 -> FX_AB_4_OU MAIS: 83 ->
FX_AB_3: 1 -> FX_AB_2: 1 -> FX_AB_1: 1
*****
Produto : CITI_MAX_ELEG_MD_TLMK
Adjacências:
CITIFINANCIA_NAO: 112761 -> FX_AB_NAO: 126874 -> CITICARD_NAO: 93468 ->
CITICARD_N_NAO: 126911 -> CITIMAX_N_NAO: 86765 -> CITICARD_ELEG_MD_TLMK: 33443 ->
CITIMAX_N_PRIMEIRA: 40146 -> CITIFINANCIA_SIM: 129475 -> FX_AB_4_OU MAIS: 156 ->
FX_AB_1: 1
*****
Produto : CITICARD_ELEG_MD_TLMK
Adjacências:
CITIFINANCIA_NAO: 40812 -> FX_AB_NAO: 48205 -> CITI_MAX_ELEG_MD_TLMK: 33443 ->
CITICARD_N_NAO: 48224 -> CITIMAX_N_NAO: 45301 -> CITI_MAX_NAO: 14781 ->
CITIFINANCIA_SIM: 24995 -> CITIMAX_N_PRIMEIRA: 2923 -> FX_AB_4_OU MAIS: 17 ->
FX_AB_3: 1 -> FX_AB_1: 1
*****
Produto : CITIMAX_N_PRIMEIRA
Adjacências:
CITIFINANCIA_NAO: 35951 -> FX_AB_NAO: 40136 -> CITI_MAX_ELEG_MD_TLMK: 40146 ->
CITICARD_NAO: 37223 -> CITICARD_N_NAO: 40146 -> CITIFINANCIA_SIM: 4195 ->
CITICARD_ELEG_MD_TLMK: 2923 -> FX_AB_4_OU MAIS: 156
*****
Produto : CITIFINANCIA_SIM
Adjacências:
FX_AB_NAO: 31734 -> CITI_MAX_ELEG_MD_TLMK: 14150 -> CITICARD_NAO: 24322 ->
CITICARD_N_NAO: 31734 -> CITIMAX_N_NAO: 27539 -> CITIMAX_N_PRIMEIRA: 4195 ->
CITI_MAX_NAO: 17584 -> CITICARD_ELEG_MD_TLMK: 7412
*****
Produto : FX_AB_4_OU MAIS
Adjacências:
CITIFINANCIA_NAO: 65 -> CITI_MAX_ELEG_MD_TLMK: 36 -> CITICARD_NAO: 48 ->
CITICARD_N_NAO: 65 -> CITIMAX_N_PRIMEIRA: 10 -> CITI_MAX_NAO: 29 -> CITIMAX_N_NAO:
55 -> CITICARD_ELEG_MD_TLMK: 17
*****
Produto : FX_AB_3
Adjacências:
CITIFINANCIA_NAO: 1 -> CITI_MAX_NAO: 1 -> CITICARD_ELEG_MD_TLMK: 1 ->
CITICARD_N_NAO: 1 -> CITIMAX_N_NAO: 1
*****
Produto : FX_AB_2
Adjacências:
CITIFINANCIA_NAO: 1 -> CITI_MAX_NAO: 1 -> CITICARD_NAO: 1 -> CITICARD_N_NAO: 1 ->
CITIMAX_N_NAO: 1
*****
Produto : FX_AB_1
Adjacências:
CITIFINANCIA_NAO: 1 -> CITI_MAX_ELEG_MD_TLMK: 1 -> CITICARD_ELEG_MD_TLMK: 1 ->
CITICARD_N_NAO: 1 -> CITIMAX_N_NAO: 1
Clique máxima pelo peso total:
- FX_AB_4_OU MAIS
- CITIMAX_N_PRIMEIRA
- CITICARD_ELEG_MD_TLMK
- CITI_MAX_ELEG_MD_TLMK
- CITIMAX_N_NAO
- CITICARD_N_NAO

```

- CITICARD_NAO - CITI_MAX_NAO - FX_AB_NAO - CITIFINANCA_NAO 2676826

Quadro 1 Relações encontradas para o caso de teste 1

Para o caso de teste 2, foram utilizadas as variáveis, presentes na base de Vendas:

- sexo
- faixa de vezes que o CPF foi abordado em uma campanha
- indicador de restrição do CPF para uma campanha
- indicador se o CPF já é de um cliente e
- estado civil

O propósito do experimento foi descobrir as relações entre tais variáveis, verificando quais características aparecem mais vezes nos registros da base Vendas.

A consulta ao banco de dados foi realizada em 65547ms (\approx 66 segundos) e a matriz de adjacência com as relações entre os valores dos atributos foi montada em 32625ms (\approx 33 segundos). Como foi realizado um filtro selecionando apenas os registros do sexo masculino e feminino (existem registro com valor de sexo “não informado” e “ignorado”, apenas 318.383 registros foram analisados).

O clique encontrado tem nós com valor 1 para faixa de abordagem (ou seja, o CPF foi abordado apenas 1 vez em campanhas de marketing), sexo feminino, sem restrições ao CPF e com indicador de que este CPF ainda não se tornou cliente da instituição. Com essas informações o analista pode inferir que as mulheres que não possuem restrição às campanhas para seus CPFs foram abordadas apenas uma vez e não responderam bem à campanha realizada, já que não se tornaram clientes. Assim, foi possível preparar mais campanhas buscando abordar novamente as pessoas que ainda não se tornaram clientes e que foram abordados apenas uma vez.

O Quadro 2 apresenta a consulta realizada sobre a base de dados, todas as relações encontradas entre as variáveis selecionadas pelo usuário e os valores encontrados como pertencentes ao clique.

Query: SELECT "IND CPF", "CITIFINANCA", "FX AB", "ESTCIVIL", "SEXO" FROM "CREDICARDCITI_CITI_VENDAS" WHERE SEXO = 'F' OR SEXO = 'M' Query execution time = ms 65547 Total de registros: 318383 Lendo os registros... Por favor, aguarde...
--

```

Cálculo das relações durou 32625ms ...
*****
Produto : IND_CPF_-_NAO
Adjacências:
CITIFINANCIA_-_NAO: 279334 -> FX_AB_-_NAO: 310155 -> ESTCIVIL_-_NAO: 310223 -> SEXO_-_F: 170677 -> SEXO_-_M: 139546 -> CITIFINANCIA_-_SIM: 30889 -> FX_AB_-_4_OU MAIS: 65 ->
FX_AB_-_3: 1 -> FX_AB_-_2: 1 -> FX_AB_-_1: 1
*****
Produto : CITIFINANCIA_-_NAO
Adjacências:
IND_CPF_-_NAO: 1070386 -> FX_AB_-_NAO: 286581 -> ESTCIVIL_-_NAO: 286649 -> SEXO_-_F: 158326 -> SEXO_-_M: 128323 -> IND_CPF_-_PROP_RECENTE: 26916 -> FX_AB_-_4_OU MAIS: 65 ->
FX_AB_-_3: 1 -> FX_AB_-_2: 1 -> FX_AB_-_1: 1
*****
Produto : FX_AB_-_NAO
Adjacências:
IND_CPF_-_NAO: 1070386 -> CITIFINANCIA_-_NAO: 731554 -> ESTCIVIL_-_NAO: 318315 ->
SEXO_-_F: 175513 -> SEXO_-_M: 142802 -> CITIFINANCIA_-_SIM: 77973 -> IND_CPF_-_
_PROP_RECENTE: 26916
*****
Produto : ESTCIVIL_-_NAO
Adjacências:
IND_CPF_-_NAO: 1070386 -> CITIFINANCIA_-_NAO: 731554 -> FX_AB_-_NAO: 493827 -> SEXO_-_F: 175553 -> SEXO_-_M: 142830 -> CITIFINANCIA_-_SIM: 77973 -> IND_CPF_-_PROP_RECENTE: 26916 -> FX_AB_-_4_OU MAIS: 103 -> FX_AB_-_3: 1 -> FX_AB_-_2: 1 -> FX_AB_-_1: 1
*****
Produto : SEXO_-_F
Adjacências:
IND_CPF_-_NAO: 1070386 -> CITIFINANCIA_-_NAO: 731554 -> FX_AB_-_NAO: 493827 ->
ESTCIVIL_-_NAO: 175553 -> CITIFINANCIA_-_SIM: 17227 -> IND_CPF_-_PROP_RECENTE: 5720 ->
FX_AB_-_4_OU MAIS: 103 -> FX_AB_-_1: 1
*****
Produto : SEXO_-_M
Adjacências:
IND_CPF_-_NAO: 139546 -> CITIFINANCIA_-_NAO: 128323 -> FX_AB_-_NAO: 142802 ->
ESTCIVIL_-_NAO: 142830 -> CITIFINANCIA_-_SIM: 77973 -> IND_CPF_-_PROP_RECENTE: 26916 ->
FX_AB_-_4_OU MAIS: 26 -> FX_AB_-_3: 1 -> FX_AB_-_2: 1
*****
Produto : CITIFINANCIA_-_SIM
Adjacências:
IND_CPF_-_NAO: 30889 -> FX_AB_-_NAO: 31734 -> ESTCIVIL_-_NAO: 31734 -> SEXO_-_M: 14507 ->
SEXO_-_F: 17227 -> IND_CPF_-_PROP_RECENTE: 5720
*****
Produto : IND_CPF_-_PROP_RECENTE
Adjacências:
CITIFINANCIA_-_NAO: 7315 -> FX_AB_-_NAO: 8160 -> ESTCIVIL_-_NAO: 8160 -> SEXO_-_M: 3284 ->
CITIFINANCIA_-_SIM: 845 -> SEXO_-_F: 4876
*****
Produto : FX_AB_-_4_OU MAIS
Adjacências:
IND_CPF_-_NAO: 65 -> CITIFINANCIA_-_NAO: 65 -> ESTCIVIL_-_NAO: 65 -> SEXO_-_F: 39 ->
SEXO_-_M: 26
*****
Produto : FX_AB_-_3
Adjacências:
IND_CPF_-_NAO: 1 -> CITIFINANCIA_-_NAO: 1 -> ESTCIVIL_-_NAO: 1 -> SEXO_-_M: 1
*****
Produto : FX_AB_-_2
Adjacências:
IND_CPF_-_NAO: 1 -> CITIFINANCIA_-_NAO: 1 -> ESTCIVIL_-_NAO: 1 -> SEXO_-_M: 1
*****
Produto : FX_AB_-_1
Adjacências:
IND_CPF_-_NAO: 1 -> CITIFINANCIA_-_NAO: 1 -> ESTCIVIL_-_NAO: 1 -> SEXO_-_F: 1
Clique máxima pelo peso total:
- FX_AB_-_1
- SEXO_-_F

```

- ESTCIVIL_-_NAO - FX_AB_-_NAO - CITIFINANCIA_-_NAO - IND_CPF_-_NAO 2140775

Quadro 2 Relações encontradas para o caso de teste 2