

# Clusterização de Empresas da Bovespa

Rafael Augusto Pedriali e Plínio Santini Dester

**Resumo**— O objetivo desse trabalho é clusterizar as empresas de capital aberto da bolsa de valores brasileira utilizando uma abordagem fundamentalista, sem envolver o preço das ações na análise, apenas os fundamentos de cada empresa, e.g., indicadores que relacionam métricas de dívida, receita, lucro, fluxo de caixa, patrimônio líquido e ativos de cada companhia. Foram utilizadas abordagens aglomerativas de clusterização, sendo selecionada a mais adequada em relação à alguns indicadores de qualidade. Os resultados obtidos auxiliam investidores a encontrarem empresas com fundamentos semelhantes sem conhecê-las plenamente, e possibilitando uma pré-seleção de empresas para compor suas carteiras de investimento.

**Palavras-Chave**— Clusterização, machine learning, Bovespa, análise fundamentalista.

## I. INTRODUÇÃO

A bolsa de valores é um ambiente de negociação de empresas com capital aberto, as quais são negociadas por meio de uma quantidade de títulos em um ambiente organizado e seguro. Quando uma empresa realiza a abertura de capital na bolsa de valores ela está “compartilhando” parte de seus direitos com as entidades que comprarem os títulos emitidos pela empresa, os quais são denominados ações. As ações são negociadas em pregões diários, e cabe aos investidores estipularem os valores de compra e venda desses papéis de acordo com o desempenho da empresa e das tendências futuras [1]. No Brasil, a empresa responsável pela bolsa de valores é a Bovespa (B3), a qual possui mais de 300 empresas listadas e é regulada pela Comissão de Valores Mobiliários (CVM), que por sua vez exige transparência na divulgação dos dados financeiros com frequência trimestral e publicamente.

Tendo em vista o atual cenário provocado pela pandemia da COVID-19 com as restrições de circulação e aglomeração de pessoas, inúmeras incertezas surgem na economia devido às mudanças comportamentais, nas rotinas, nos trabalhos e nos hábitos de consumo das pessoas. Estas mudanças influenciam nas incertezas com relação ao futuro das atividades sociais e econômicas, afetando a relação entre empresas e investidores. Foi observada uma desvalorização sistêmica expressiva na bolsa de valores devido às incertezas da pandemia, mas no decorrer do ano, com a divulgação dos balanços trimestrais, ocorreu a recuperação das empresas. No entanto, nem todas voltaram ao período pré-pandemia mesmo tendo apresentado estabilidade financeira. Tais resultados aliados com a queda da taxa básica de juros (Selic) próximo ao patamar de 2% têm atraído um público novo de pequenos e médios investidores [2]. De fato, do fim de 2018 até o fim de 2020 o número de pessoas físicas na bolsa triplicou [3].

Rafael A. Pedriali e Plínio S. Dester estão no *The Wireless Technology Laboratory* (WissTek), Faculdade de Engenharia Elétrica e de Computação, Universidade Estadual de Campinas (e-mail: rafaug.p@gmail.com; pliniodes-ter@gmail.com). Este trabalho foi financiado em parte pelo CNPq código 141425/2019-9 e FAPESP código 2017/21347-0.

Este trabalho tem o objetivo de usar os indicadores financeiros das empresas para realizar um processo de clusterização não supervisionada dos dados, resultando em uma organização pertinente e adequada das empresas. Existem diversos trabalhos que realizam a clusterização de empresas, alguns deles são dados a seguir. O artigo [4] emprega uma abordagem hierárquica de clusterização de ações utilizando o preço de fechamento diário, e verifica que isso pode ser usado para agrupar as empresas de acordo com uma classificação setorial próxima à aceita pela comunidade de investidores. Em [5], o autor realiza a clusterização sobre um conjunto de empresas listadas na Bovespa para agrupá-las usando estatísticas sobre o preço de fechamento diário e compara o resultado obtido com a classificação setorial oficial da Bovespa. Em [6] é utilizado um mapa auto-organizável para realizar a clusterização das empresas a partir dos indicadores financeiros como liquidez corrente, dívida líquida sobre patrimônio líquido, *dividend yield*, EBIT, ROE, ROA, etc. O artigo [7] realiza a clusterização de empresas iranianas utilizando o algoritmo *k-means* para agrupar-las através dos dados financeiros das empresas. Em [8] a clusterização é feita através do mapa auto-organizável, *k-means* e *Fuzzy C-means*, com o propósito de montar automaticamente uma carteira de investimentos na bolsa de valores da Índia, os atributos usados na clusterização foram indicadores financeiros que dependem do preço das ações. As principais diferenças entre o presente trabalho e [6, 7, 8] é que este foi usado a abordagem aglomerativa para clusterização em um conjunto diferente de atributos (indicadores financeiros) que avaliam o equilíbrio financeiro das empresas sem depender do valor de mercado das companhias. Desta forma, este trabalho não tem a intenção de envolver o preço dos ativos e derivativos. De fato, não foi utilizado nenhum atributo que depende do valor de mercado das empresas, pois adotamos a hipótese do mercado eficiente [9], e assim não há ineficiências do mercado a serem exploradas nesse sentido.

Dentre os indicadores financeiros disponíveis, foram utilizados: patrimônio líquido, receita líquida, EBIT, lucro líquido, dívida bruta, fluxo de caixa, CAPEX, composição dos ativos e dividendos pagos. O método desenvolvido proporciona uma forma mais simples de detectar resultados financeiros semelhantes. Agrupando, por exemplo, empresas alavancadas, empresas com margem líquida baixa, empresas com dificuldade em pagar suas dívidas, empresas que exigem um alto investimento para manter suas operações, e assim por diante. Nesse sentido, espera-se que esse trabalho auxilie o investidor à ter uma visão geral de como as empresas estão distribuídas na bolsa de valores, e em selecionar as empresas que, para o investidor, faz sentido realizar um estudo mais aprofundado.

Este artigo é estruturado da seguinte forma. A Seção II contém informações preliminares para o desenvolvimento da pesquisa: posicionamento da pesquisa, base conceitual e me-

todologia utilizada. A Seção III apresenta os resultados e a análise dos resultados. Finalmente, a Seção IV conclui o artigo e apresenta possíveis trabalhos futuros.

## II. PRELIMINARES

### A. Posicionamento da Pesquisa

Considerando o conjunto de dados financeiros das empresas de capital aberto da Bovespa, deseja-se separá-las em diferentes *clusters*, cada um englobando certas características financeiras das empresas, limitando-se à uma análise puramente fundamentalista. Até onde sabemos, não existe nenhum estudo publicado nesse sentido para empresas brasileiras. Diante disso, utilizar técnicas de *Machine Learning* é uma estratégia pertinente para realizar o agrupamento das empresas. Existem vários métodos de clusterização presentes na literatura [10], dentre eles temos o *density-based spatial clustering of applications with noise* (DBSCAN), o *k-means* e a clusterização hierárquica. Este artigo implementa a clusterização hierárquica junto com a normalização dos dados e a atribuição de pesos nos indicadores financeiros mais relevantes da abordagem utilizada, a qual está detalhada na Seção II-B.

### B. Base Conceitual

A clusterização hierárquica com abordagem aglomerativa é feita iniciando cada amostra do conjunto de dados como um *cluster*. Em seguida, os *clusters* de menor distância são aglomerados até que forma-se um único *cluster* com todo o conjunto de dados. Escolhe-se ao fim do processo, o momento da clusterização em que tem-se uma quantidade razoável de *clusters* para o propósito do problema [10].

O que diferencia os métodos de clusterização aglomerativa é a definição da distância entre dois *clusters*. Os métodos explorados nesse trabalho são definidos a seguir. Sejam  $\mathcal{U}$  e  $\mathcal{V}$  dois *clusters* de dados, então definimos [11]

#### Single-linkage:

$$d_S(\mathcal{U}, \mathcal{V}) \triangleq \min_{u \in \mathcal{U}, v \in \mathcal{V}} d(u, v), \quad (1)$$

#### Average-linkage:

$$d_A(\mathcal{U}, \mathcal{V}) \triangleq \sum_{u \in \mathcal{U}, v \in \mathcal{V}} \frac{d(u, v)}{|\mathcal{U}| |\mathcal{V}|}, \quad (2)$$

#### Complete-linkage:

$$d_C(\mathcal{U}, \mathcal{V}) \triangleq \max_{u \in \mathcal{U}, v \in \mathcal{V}} d(u, v), \quad (3)$$

#### Ward-linkage:

$$d_W(\mathcal{U}, \mathcal{V})^2 \triangleq \frac{1}{|\mathcal{V}| + |\mathcal{U}|} \left( (|\mathcal{V}| + |\mathcal{S}|) d_W(\mathcal{V}, \mathcal{S})^2 + (|\mathcal{V}| + |\mathcal{T}|) d_W(\mathcal{V}, \mathcal{T})^2 - |\mathcal{V}| d_W(\mathcal{S}, \mathcal{T})^2 \right), \quad (4)$$

onde  $\mathcal{S}$  e  $\mathcal{T}$  são os *clusters* que formaram  $\mathcal{U}$  anteriormente, i.e.,  $\mathcal{S} \cup \mathcal{T} = \mathcal{U}$ . A Eq. (4) trata-se de uma recorrência e o caso base é quando  $\mathcal{U}$  e  $\mathcal{V}$  possuem um único elemento e, nesse momento, usa-se a definição escolhida para distância entre vetores. O operador  $|\cdot|$  denota o número de elementos do *cluster*.

Note que para todas as definições é necessário escolher alguma métrica de distância para vetores. No presente trabalho foi utilizada a distância euclidiana, ou seja,

$$d(u, v) = \sqrt{\sum_{i=1}^n (u_i - v_i)^2}, \quad u, v \in \mathbb{R}^n, \quad (5)$$

onde  $u_i, v_i$  são as  $i$ -ésimas componentes dos vetores  $u, v$ , respectivamente.

Também existem indicadores para medir a qualidade das clusterizações as quais são muito importantes para validar a eficiência dos agrupamentos de dados. A seguir são apresentadas três delas, o *Davies-Bouldin score*, *Silhouette Coefficient score* e o *Calinski-Harabasz score* [12, Seção 2.3.10].

#### Davies-Bouldin score:

Este índice mede a similaridade média entre os *clusters*, ou seja, a distância entre diferentes *clusters* são comparadas com o tamanho de seus próprios *clusters*. Seu índice é dado por

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{i \neq j} \frac{s_i + s_j}{d_{ij}}, \quad (6)$$

onde  $k$  é o número total de *clusters*,  $s_i$  é a média das distâncias entre cada ponto do  $i$ -ésimo *cluster* com o centroide de seu respectivo *cluster*, e  $d_{ij}$  é a distância entre os centroides dos *clusters*  $i$  e  $j$ . Quanto menor seu valor mais bem organizado estão os *clusters*.

#### Silhouette Coefficient score:

Este indicador é calculado com a seguinte expressão,

$$SC = \sum_{i=1}^k S_i, \quad (7)$$

em que  $S_i$  é o *Silhouette Coefficient* para o  $i$ -ésimo *cluster* e expresso por  $S_i = \frac{b_i - a_i}{\max(a_i, b_i)}$ ,  $a_i$  representa a distância média entre uma amostra e todos os outros pontos do  $i$ -ésimo *cluster* e  $b_i$  é a distância média entre uma amostra do  $i$ -ésimo *cluster* e todos os pontos do outro *cluster* mais próximo. Este indicador está contido no intervalo  $[-1, 1]$ , o melhor agrupamento corresponde ao valor 1, o pior ao valor  $-1$ , e o valor 0 corresponde à sobreposição de *clusters*.

#### Calinski-Harabasz score:

Este indicador também é conhecido como critério da razão de variância, ou seja, corresponde à razão entre a soma da dispersão entre os *clusters* e a dispersão interna ao *cluster* para todos os *clusters*, e sua expressão é definida por

$$CH = \frac{|\mathcal{E}| - k}{k - 1} \frac{\text{Tr} \left( \sum_{i=1}^k |\mathcal{V}| (c_{\mathcal{V}} - c_{\mathcal{E}})(c_{\mathcal{V}} - c_{\mathcal{E}})^T \right)}{\text{Tr} \left( \sum_{i=1}^k \sum_{x \in \mathcal{V}} (x - c_{\mathcal{V}})(x - c_{\mathcal{V}})^T \right)} \quad (8)$$

em que  $\text{Tr}(\cdot)$  é o operador traço de uma matriz,  $\mathcal{E}$  é o tamanho do conjunto total de dados,  $c_{\mathcal{E}}$  e  $c_{\mathcal{V}}$  correspondem ao centroides de  $\mathcal{E}$  e  $\mathcal{V}$  respectivamente. Quanto maior o valor do indicador, mais organizados estão os dados.

### C. Metodologia

Existem diversos indicadores financeiros para identificar as características das empresas, como o lucro líquido, EBIT, patrimônio líquido, dívida, ativo total, entre outros. A maioria destes indicadores não são normalizados, o que dificulta a

clusterização das empresas. Por exemplo, a Petrobras S.A. tem valor de mercado cem vezes maior que a Camil S.A., o que torna as grandezas mencionadas incomparáveis entre as duas empresas. Dessa forma, seria interessante selecionar indicadores que já são naturalmente normalizados para realizar a clusterização. Os indicadores escolhidos na clusterização foram: liquidez corrente, dívida líquida/EBITDA, CAPEX/FCO, FCL/RL, ROA, ROIC, margem EBITDA, margem bruta, PL/AT, RL/AT, *Dividend Payout* e Intan/AT. Abaixo uma breve explicação de cada um deles [1] e com a escala utilizada indicada entre parênteses,

- **Liquidez corrente** (0;2): Razão entre ativo circulante e passivo circulante. Mostra a capacidade da empresa em pagar suas dívidas de curto prazo (quanto maior, melhor). Esse indicador foi saturado em 2, pois acima disso a empresa é capaz de pagar suas dívidas de curto prazo com muita folga e, além desse valor, não interessa mais para a análise.
- **Dívida líquida/EBITDA** (0;6): Razão entre a dívida bruta menos as disponibilidades e o EBITDA<sup>1</sup>. Traduz a capacidade da empresa em pagar suas dívidas (quanto menor, melhor). Para a clusterização, esse indicador foi saturado entre 0 e 6, uma vez que empresas com dívida líquida menor que 0 não tem dívida. Além disso, existem empresas com EBITDA muito baixo, e esse indicador fica alto, o que prejudicaria a clusterização, por isso a saturação em 6.
- **CAPEX/FCO** (0;1,25): razão entre gastos com investimento e manutenção dos ativos essenciais para empresa e o fluxo de caixa operacional. Reflete o quanto a empresa gasta para manter e investir em seus equipamentos e bens. Esse atributo foi saturado em 1,25, pois mais que isso indica casos excepcionais.
- **FCL/RL** (−0,25;0,5): razão entre o fluxo de caixa livre<sup>2</sup> e a receita líquida<sup>3</sup>. Representa a proporção da receita que poderia ser extraída da empresa em forma de dinheiro sem prejudicar suas operações. Seu valor foi limitado a −0,25 e saturado em 0,5 devido aos *outliers*.
- **ROA** (−0,1;0,2): razão entre o lucro líquido e o ativo total<sup>4</sup>. É a capacidade da empresa em usar seus ativos para gerar lucro. Os valores são limitados a −0,1 e saturados em 0,2.
- **ROIC** (−0,1;0,25): razão entre o lucro operacional e o capital investido. Seu resultado avalia a eficiência da empresa em gerar retorno com o uso de seu capital. Novamente, os valores foram limitados a −0,1 e saturados em 0,25.
- **Margem EBITDA** (−0,25;1): razão entre o EBITDA e a receita líquida. Mostra a porcentagem da receita líquida que sobra após descontar os custos inerentes à operação da empresa. Este indicador está limitado em −0,25, pois a parte negativa é suficiente para informar que a empresa opera com prejuízo.
- **Margem bruta** (−1;1): razão entre o lucro bruto<sup>5</sup> e a

receita líquida. Quanto maior esse indicador, maior é a capacidade de aumentar o lucro com o aumento de escala.

- **PL/AT** (−0,3;2): razão entre o patrimônio líquido<sup>6</sup> e o ativo total. O valor foi limitado a −0,3 e saturado em 2 a fim de evitar os efeitos dos *outliers*.
- **RL/AT** (0;2): razão entre a receita líquida e o ativo total. Representa a capacidade da empresa em usar seus ativos para gerar receita líquida.
- **Intan/AT** (0;1): razão entre ativo intangível<sup>7</sup> e ativo total.
- **Dividend Payout** (0;1): Porcentagem do lucro líquido que é distribuído para os acionistas. Mostra se a empresa está reinvestindo os lucros ou distribuindo-os para os sócios.

Esses indicadores são normalizados para todas as empresas com o intuito de manter pesos semelhantes para cada atributo. Para os atributos mais importantes, o peso é multiplicado por dois. No caso escolhemos para ter um peso maior o principal indicador de endividamento “Dívida Líq./Ebitda”, o indicador que mostra se a empresa gera muita receita “RL/AT” e um dos principais indicadores do lucro operacional da empresa “Margem Ebitda”. Ademais, é utilizada a média desses indicadores dos últimos três anos (2018, 2019 e 2020) com o intuito de amenizar eventos não recorrentes. Foram usadas 151 empresas, retirou-se os bancos devido à grande diferença na estrutura financeira, e as empresas que fizeram IPO recentemente, pois podem apresentar distorções nos indicadores devido à injeção de capital.

### III. ANÁLISE E RESULTADOS

Primeiramente, foram analisadas as seguintes abordagens aglomerativas usando a ferramenta Scikit-learn [13] implementadas em Python: *single-linkage*, *average-linkage*, *complete-linkage* e *ward-linkage*, que são definidas pelas equações (1), (2), (3) e (4), respectivamente. Em seguida, variou-se o número de *clusters* para cada método e mediu-se os indicadores de qualidade Davies-Bouldin score, *Silhouette Coefficient score* e Calinski-Harabasz score

A Fig. 1 apresenta os três indicadores de qualidade em clusterizações, em cada gráfico são comparados os *scores* para cada método aglomerativo. Esses indicadores mostram informações para auxiliar na escolha do método mais eficiente de clusterização. De modo geral, nota-se que o método *average-linkage* é o mais promissor para um número mais elevado de *clusters*. Porém, ao realizarmos a clusterização com o *average-linkage*, apareceram muitos *clusters* com um único elemento e um *cluster* com muitos elementos. O que, para nosso propósito, não é interessante. Por outro lado, o método *ward-linkage* se destacou nos *scores Silhouette Coefficient* e Calinski-Harabasz score, porém falhou no Davies-Bouldin score com o pior resultado. Considerando os indicadores de qualidade apresentados, a melhor clusterização foi a *ward-linkage*, a qual se destacou em dois dos três indicadores de performance e ainda possui uma equilibrada distribuição de elementos nos *clusters*. Dessa forma, o método *ward-linkage* foi o escolhido para realizar a clusterização final.

<sup>6</sup>Patrimônio Líquido (PL) é o valor total que os sócios ou acionistas possuem na empresa.

<sup>7</sup>Ativo intangível é um ativo não monetário identificável sem substância física, por exemplo marcas, patentes, direitos autorais, *software*, clientes, etc.

<sup>1</sup>EBITDA é o lucro antes de juros, impostos, depreciação e amortização.

<sup>2</sup>Fluxo de caixa livre (FCL) é a soma do caixa proveniente das atividades operacionais menos os gastos com investimentos.

<sup>3</sup>Receita líquida (RL) É a receita gerada pelas atividades da empresa menos deduções (devoluções, descontos e imposto sobre vendas).

<sup>4</sup>Ativo total (AT) é o conjunto de todos os bens, valores, créditos e direitos pertencentes à empresa.

<sup>5</sup>Lucro bruto é a receita líquida menos os custos de produção.



Fig. 1. Indicadores de qualidade da clusterização em função do número de clusters para vários métodos.

Além disso, foi selecionado 10 clusters como a quantidade ideal para este trabalho, pois mais que essa quantidade formariam clusters muito pequenos e ficaria complexo discernir as principais características de cada cluster. Por outro lado, menos que isso formariam clusters muito grandes e perderia o propósito de auxiliar o investidor.

A Fig. 2 mostra a distribuição dos objetos (empresas) em seus clusters no espaço de atributos, considerando a melhor clusterização encontrada. É importante destacar que o espaço de atributos tem 12 dimensões e não é possível representá-lo no plano do papel, a alternativa é apresentá-lo em 6 gráficos com 2 dimensões cada. A Tabela I detalha a classificação dos clusters apresentando as empresas e a descrição das características predominantes em cada cluster. Para o algoritmo proposto, a aplicabilidade prática é ilustrada na Tabela I, onde as empresas são agrupadas por seus fundamentos, e são identificadas e descritas as características mais predominantes de cada cluster.

Apesar da clusterização ter sido satisfatória para a maioria das empresas, é esperado que algumas delas não se adéquem à nenhum dos cluster. Por exemplo, a GSHP é uma empresa muito endividada, possui patrimônio líquido negativo e deu prejuízo nos três últimos anos, dessa forma, espera-se que ela seja posicionada no cluster 1, porém devido à alta margem EBITDA e ao controle das dívidas de curto prazo, ela foi posicionada no cluster 9.

#### IV. CONCLUSÕES E PERSPECTIVAS FUTURAS

Esse projeto apresenta uma abordagem de clusterização hierárquica das empresas de capital aberto da Bovespa com foco em parâmetros fundamentalistas. Após inúmeros testes, foram selecionados os seguintes resultados financeiros para clusterização: liquidez corrente, dívida líquida/EBITDA, CAPEX/FCO, FCL/RL, ROA, ROIC, margem EBITDA, margem bruta, PL/AT, RL/AT, Dividend Payout e Intangível/AT. Diversos métodos de clusterização hierárquica foram comparados por meio dos indicadores de qualidade *Silhouette coefficient score*, *Davies-Bouldin score* e *Calinski-Harabasz score*. O método que mais se adequou para realizar a clusterização foi o *ward-linkage*, o qual apresentou indicadores razoáveis e realizou agrupamentos equilibrados e que faziam sentido do ponto de vista fundamentalista. Além disso, pretende-se adicionar outros parâmetros fundamentalistas, comparar com outras classificações usadas pelos investidores, e divulgar

o código no GitHub. Dessa forma, pessoas interessadas no mercado de ações podem utilizá-lo com a finalidade de auxiliar em suas decisões de investimento.

#### REFERÊNCIAS

- [1] H. Santiago. *Análise Fundamentalista de Empresas para Investimentos em Ações: Capital e Valor*. Capital e Valor, 2015.
- [2] L. Quintino e F. Mendes. “Por que a onda de novatos na bolsa pode gerar um ciclo virtuoso no Brasil”. Em: *VEJA* n. 2702 (set. de 2020). ISSN: 0100-7122.
- [3] F. Ferreira. *No ano, número de investidores pessoas físicas na Bolsa cresce mais de +80%*. Rel. técn. XP Investimentos CCTVM S/A, nov. de 2020.
- [4] K. A. J. Doherty et al. “Hierarchical topological clustering learns stock market sectors”. Em: *2005 ICSC Congress on Computational Intelligence Methods and Applications*. IEEE, 2005, p. 6.
- [5] R. A. Torres e H. C. V. Lopes. “Aplicação de métodos de clusterização em um estudo sobre o mercado acionário brasileiro”. Diss. de mestr. PUC-RIO, 2014.
- [6] S. S. H. Kelvin. *Using Self-Organizing Maps (SOM) to Cluster Stocks and Financial Ratios*. 2006.
- [7] M. Momeni, M. Mohseni e M. Soofi. “Clustering stock market companies via k-means algorithm”. Em: *Kuwait Chapter of Arabian Journal of Business and Management Review* 33.2578 (2015), pp. 1–10.
- [8] S.R. Nanda, B. Mahanty e M.K. Tiwari. “Clustering Indian stock market data for portfolio management”. Em: *Expert Systems with Applications* 37.12 (2010), pp. 8793–8798. ISSN: 0957-4174.
- [9] Eugene F. Fama. “Efficient Capital Markets: A Review of Theory and Empirical Work”. Em: *The Journal of Finance* 25.2 (1970), pp. 383–417.
- [10] C.C. Aggarwal. *Machine Learning for Text*. Springer International Publishing, 2018. ISBN: 9783319735313.
- [11] J. Webber. *SciPy: Open source scientific tools for Python*. 2014.
- [12] *User guide: contents*. scikit-learn developers. 2020.
- [13] F. Pedregosa et al. “Scikit-learn: Machine Learning in Python”. Em: *Journal of Machine Learning Research* 12 (2011), pp. 2825–2830.

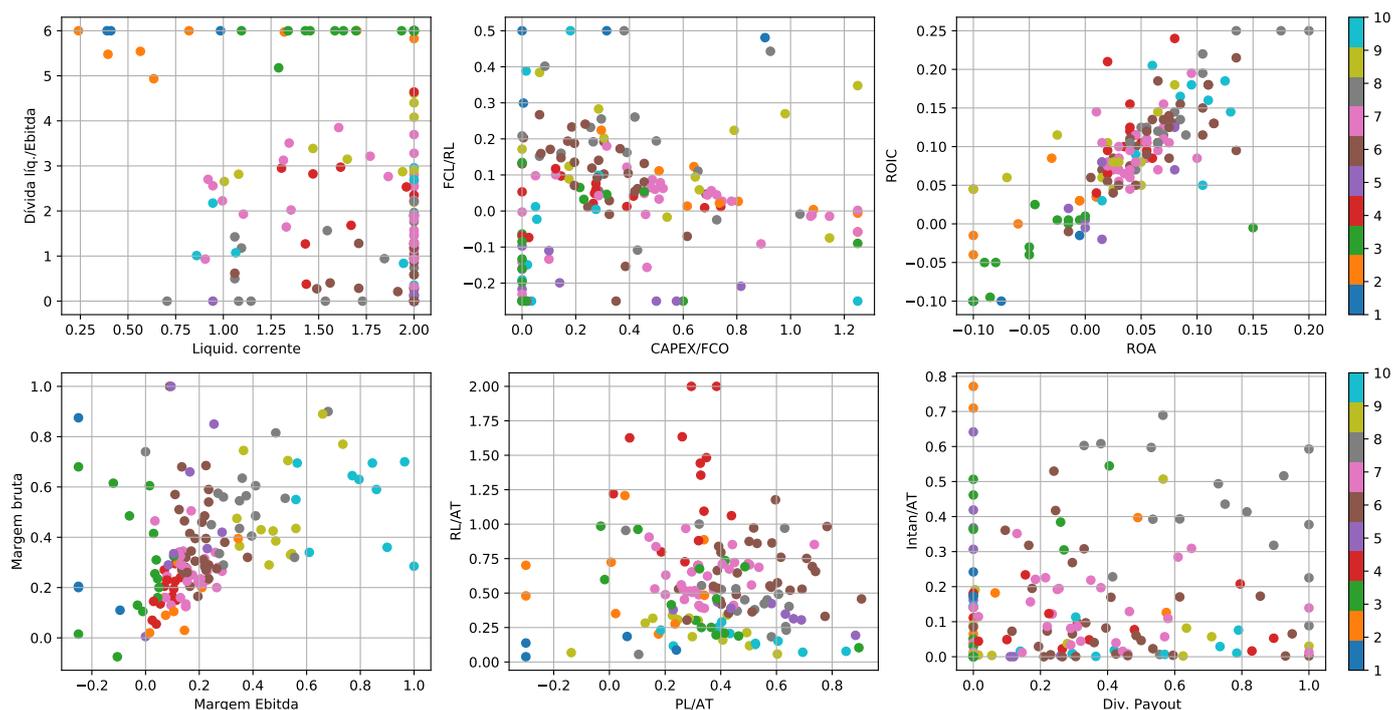


Fig. 2. Clusterização escolhida no espaço de atributos. Cada cor representa um *cluster*.

<i>cluster</i>	<i>empresas</i>	<i>descrição</i>
1	CVCB, RSID, PDGR, LUPA	dívidas de curto e longo prazo descontroladas, prejuízo elevado, patrimônio líquido baixo ou negativo, receita baixa
2	GOLL, AZUL, BRKM, PMAM, DASA, OIBR, ECOR, MYPK, TPIS	dívida descontrolada, alto CAPEX, margens apertadas, patrimônio líquido baixo, significativa parte dos ativos são intangíveis
3	EMBR, VVAR, BKBR, AMAR, LLI, VLID, HBOR, SGPS, RDNI, GFSA, BBRK, FHER, CRDE, LIQO, COGN	dívida descontrolada, baixo CAPEX, consome caixa, tem prejuízo
4	JBSS, CRFB, MGLU, BRDT, RADL, BRFS, SULA, UGPA, MRFG, CAML, BEEF, PFRM, POSI	dívida controlada, margens apertadas, receita líquida elevada, paga dividendos
5	IRBR, BTOW, LINX, EZTC, TRIS, LPSB, SQIA, APER	sem dívida, gera pouco caixa, margens apertadas, patrimônio líquido significativo, não paga dividendos
6	VALE, WEGE, LREN, HAPV, NATU, GNDI, PSSA, MDIA, ALPA, USIM, DTEX, CYRE, TOTS, GRND, HGTX, ARZZ, UNIP, LEVE, PARD, TEND, FESA, TGMA, EVEN, DIRR, SLCE, MILS, GUAR, CGRA, CARD, UCAS, MTSA	sem dívida, gera caixa, tem lucro, patrimônio líquido significativo, paga dividendos
7	CPFE, RENT, LAME, GGBR, PCAR, CSNA, CMIG, CSAN, EQTL, REDE, CPLE, ENBR, MRVE, GOAU, LCAM, COCE, LIGT, MOVI, POMO, RAPT, TUPY, CNTO, CLSC, OFSA, VULC, FRAS, ROMI, SHUL, PTBL, LOGN, ETER	dívida controlada, alto CAPEX, gera pouco caixa, tem lucro, paga dividendos
8	ABEV, VIVT, B3SA, SBSP, CIEL, HYPE, CGAS, ODPV, CSMG, SAPR, FLRY, QUAL, ENAT, SEER, WIZS, YDUQ, TIMS	sem dívida, gera caixa, margens altas, significativa parte dos ativos são intangíveis, paga muitos dividendos
9	PETR, ELET, SUZB, EGIE, RAIL, CCRO, KLBN, BRML, ENEV, SMTO, WSON, SCAR, GSHP	dívida controlada, alto CAPEX, margens altas, paga dividendos
10	ITSA, TAEE, CESP, IGTA, ALUP, AESB, TRPL, ALSO, PRIO, JHSF	dívida controlada, gera muito caixa, margens altas, paga dividendos, possui poucos ativos intangíveis

TABELA I  
Clusters E SUAS CARACTERÍSTICAS.