Setores econômicos em um espaço ultramétrico

Suponhamos que temos os preços de fechamento de várias ações em $N+1$ pregões. É conveniente definirmos a seguinte variável:

$$S_i\left(t\right) \quad = \quad \ln\left[\frac{Y_i\left(t\right)}{Y_i\left(t-1\right)}\right],$$

onde t indexa os pregões variando de 0 a N e Y_i é o preço da i-ésima ação. Ao invés de indexarmos o pregão por (t), como acima, escrevamos assim:

$$S_{i\alpha} \quad = \quad \ln\left[Y_{i\alpha}\right] - \ln\left[Y_{i(\alpha-1)}\right],$$

onde agora índices gregos denotam os números dos pregões. Dessa forma, acima, α varia de 1 a N. Temos, portanto, N valores de S_i.

É conveniente definirmos:

$$U_{i\alpha} \quad = \quad \frac{1}{\sqrt{N}}\frac{S_{i\alpha}-\langle S_i\rangle}{\sqrt{\langle S_i^2\rangle-\langle S_i\rangle^2}},$$

onde

$$\langle S_i\rangle \quad = \quad \frac{1}{N}\sum_{\alpha=1}^{N} S_{i\alpha}$$

e

$$\left\langle S_i^2\right\rangle \quad = \quad \frac{1}{N}\sum_{\alpha=1}^{N} S_{i\alpha}^2.$$

Podemos construir, para cada ação, um vetor no espaço euclidiano de N dimensões com as componentes $U_{i\alpha}$. O produto escalar desse vetor para duas ações dá:

$$\mathbf{U}_i\cdot\mathbf{U}_j \quad = \quad \sum_{\alpha=1}^{N} U_{i\alpha}U_{j\alpha} = \frac{1}{N}\sum_{\alpha=1}^{N}\frac{S_{i\alpha}-\langle S_i\rangle}{\sqrt{\langle S_i^2\rangle-\langle S_i\rangle^2}}\frac{S_{j\alpha}-\langle S_j\rangle}{\sqrt{\langle S_j^2\rangle-\langle S_j\rangle^2}},$$

isto é,

$$\mathbf{U}_i\cdot\mathbf{U}_j \quad = \quad \frac{\frac{1}{N}\sum_{\alpha=1}^{N}\left(S_{i\alpha}-\langle S_i\rangle\right)\left(S_{j\alpha}-\langle S_j\rangle\right)}{\sqrt{\left(\langle S_i^2\rangle-\langle S_i\rangle^2\right)\left(\langle S_j^2\rangle-\langle S_j\rangle^2\right)}},$$

ou seja,

$$\mathbf{U}_i\cdot\mathbf{U}_j \quad = \quad \frac{\langle\left(S_i-\langle S_i\rangle\right)\left(S_j-\langle S_j\rangle\right)\rangle}{\sqrt{\left(\langle S_i^2\rangle-\langle S_i\rangle^2\right)\left(\langle S_j^2\rangle-\langle S_j\rangle^2\right)}},$$

ou ainda,

$$\mathbf{U}_i \cdot \mathbf{U}_j = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\left\langle S_i^2 - \langle S_i \rangle^2 \right\rangle \left\langle S_j^2 - \langle S_j \rangle^2 \right\rangle}}.$$

O coeficiente de correlação, ρ_{ij}, entre a i-ésima e a j-ésima ações é, convenientemente, definido por

$$\rho_{ij} = \frac{\langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle}{\sqrt{\left\langle S_i^2 - \langle S_i \rangle^2 \right\rangle \left\langle S_j^2 - \langle S_j \rangle^2 \right\rangle}}$$

e, portanto,

$$\mathbf{U}_i \cdot \mathbf{U}_j = \rho_{ij}.$$

Esses vetores têm, portanto, módulo unitário:

$$\mathbf{U}_i \cdot \mathbf{U}_i = \rho_{ii} = 1.$$

É fácil ver que

$$-1 \leqslant \rho_{ij} \leqslant 1$$

se notarmos que

$$\begin{aligned} \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle) \rangle &= \langle S_i S_j - S_i \langle S_j \rangle - \langle S_i \rangle S_j + \langle S_i \rangle \langle S_j \rangle \rangle \\ &= \langle S_i S_j \rangle - \langle S_i \rangle \langle S_j \rangle \end{aligned}$$

e

$$\sqrt{\left\langle S_i^2 - \langle S_i \rangle^2 \right\rangle \left\langle S_j^2 - \langle S_j \rangle^2 \right\rangle} = \sqrt{\left\langle (S_i - \langle S_i \rangle)^2 \right\rangle \left\langle (S_j - \langle S_j \rangle)^2 \right\rangle}.$$

Então, obviamente,

$$\left\langle \left[(S_i - \langle S_i \rangle) - \lambda (S_j - \langle S_j \rangle) \right]^2 \right\rangle \geqslant 0$$

para todo λ real. Logo,

$$\begin{aligned} \left\langle \left[(S_i - \langle S_i \rangle) + \lambda (S_j - \langle S_j \rangle) \right]^2 \right\rangle &= \left\langle (S_i - \langle S_i \rangle)^2 \right\rangle \\ &- 2\lambda \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle) \rangle + \lambda^2 \left\langle (S_j - \langle S_j \rangle)^2 \right\rangle, \end{aligned}$$

ou seja,

$$2\lambda \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle) \rangle \leqslant \left\langle (S_i - \langle S_i \rangle)^2 \right\rangle + \lambda^2 \left\langle (S_j - \langle S_j \rangle)^2 \right\rangle.$$

Como essa desigualdade deve valer para todo λ real, escolhemos

$$\lambda \quad = \quad \frac{\langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle}{\left\langle (S_j - \langle S_j \rangle)^2 \right\rangle}$$

e obtemos

$$\begin{aligned} 2\frac{\langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle}{\left\langle (S_j - \langle S_j \rangle)^2 \right\rangle} \langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle \quad &\leqslant \quad \left\langle (S_i - \langle S_i \rangle)^2 \right\rangle \\ &+ \quad \frac{\langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle^2}{\left\langle (S_j - \langle S_j \rangle)^2 \right\rangle}, \end{aligned}$$

isto é,

$$\langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle^2 \quad \leqslant \quad \left\langle (S_i - \langle S_i \rangle)^2 \right\rangle \left\langle (S_j - \langle S_j \rangle)^2 \right\rangle$$

e, portanto,

$$\frac{|\langle (S_i - \langle S_i \rangle)(S_j - \langle S_j \rangle)\rangle|}{\sqrt{\left\langle (S_i - \langle S_i \rangle)^2 \right\rangle \left\langle (S_j - \langle S_j \rangle)^2 \right\rangle}} \quad \leqslant \quad 1,$$

demonstrando que

$$|\rho_{ij}| \quad \leqslant \quad 1,$$

isto é,

$$-1 \leqslant \quad \rho_{ij} \quad \leqslant 1.$$

A distância euclidiana entre dois desses vetores é obtida como segue:

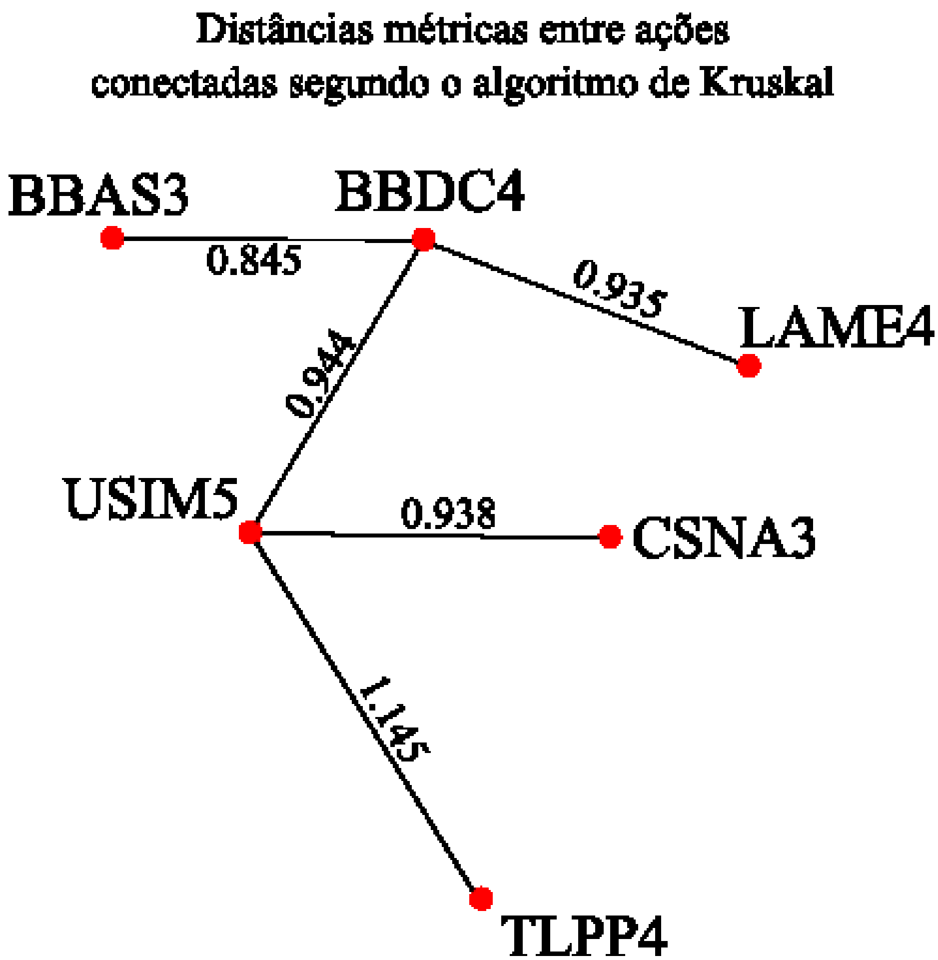
$$\begin{aligned} d_{ij} \quad &= \quad \sqrt{(\mathbf{U}_i - \mathbf{U}_j) \cdot (\mathbf{U}_i - \mathbf{U}_j)} \\ &= \quad \sqrt{\mathbf{U}_i \cdot (\mathbf{U}_i - \mathbf{U}_j) - \mathbf{U}_j \cdot (\mathbf{U}_i - \mathbf{U}_j)} \\ &= \quad \sqrt{\mathbf{U}_i \cdot \mathbf{U}_i - \mathbf{U}_i \cdot \mathbf{U}_j - \mathbf{U}_j \cdot \mathbf{U}_i + \mathbf{U}_j \cdot \mathbf{U}_j} \\ &= \quad \sqrt{1 - \rho_{ij} - \rho_{ji} + 1} \\ &= \quad \sqrt{2\,(1 - \rho_{ij})}. \end{aligned}$$

Com base nessa distância, podemos definir uma outra, chamada distância ultramétrica, de acordo com o algoritmo de Kruskal. A idéia é construir uma árvore de extensão mínima. É fácil depois de um pouco de treinamento. Tudo começa com a idéia de que o espaço métrico de N dimensões acima, com a distância d_{ij}, pode ser utilizado para gerar um espaço ultramétrico conforme o algoritmo de Kruskal.

Usando dados que acumulei desde meados de 2004 até meados de 2009, referentes a todos os pregões desse período, selecionei os preços de fechamento das ações dos Bancos Bradesco (BBDC4) e do Brasil (BBAS3), da Usiminas (USIM5), da Companhia Siderúrgica Nacional (CSNA3), das Lojas Americanas (LAME4) e da TELESP (TLPP4). Com esses números, construí a seguinte tabela de distâncias d_{ij} :

d_{ij} : $\,^{j\rightarrow}_{i\downarrow}$	BBDC4	BBAS3	USIM5	CSNA3	LAME4	TLPP4
BBDC4	0	0.845	0.944	1.062	0.935	1.160
BBAS3	0.845	0	0.977	1.084	0.989	1.172
USIM5	0.944	0.977	0	0.938	1.026	1.145
CSNA3	1.062	1.084	0.938	0	1.095	1.227
LAME4	0.935	0.989	1.026	1.095	0	1.160
TLPP4	1.160	1.172	1.145	1.227	1.160	0

Vemos que a menor distância é 0.845, entre BBAS3 e BBDC4, dois bancos. A seguinte menor distância é 0.935, entre BBDC4 e LAME4, um banco e uma empresa do setor de consumo geral. Depois, é 0.938, entre USIM5 e CSNA3, duas empresas do setor de siderurgia e metalurgia. Enfim, essas distâncias são menores entre empresas de mesmo setor ou de setores interdependentes. Continuando dessa forma e deixando de desenhar linhas que conectam ações de modo a dar origem a um circuito fechado, obtemos a figura abaixo para a árvore de extensão mínima dessas seis ações.



Agora precisamos construir uma hierarquia de distâncias subdominantes ultramétricas, analogamente ao que está feito no livro de Rosario Nunzio Mantegna e de H. Eugene Stanley. A figura abaixo mostra a hierarquia obtida entre as ações que estamos considerando. Com base nessa hierarquia, a tabela abaixo mostra as distâncias ultramétricas $\hat{d_{ij}}$.

$\hat{d_{ij}}$: $\overset{j\rightarrow}{i\downarrow}$	BBDC4	BBAS3	USIM5	CSNA3	LAME4	TLPP4
BBDC4	0	0.845	0.944	0.944	0.935	1.145
BBAS3	0.845	0	0.944	0.944	0.935	1.145
USIM5	0.944	0.944	0	0.938	0.944	1.145
CSNA3	0.944	0.944	0.938	0	0.944	1.145
LAME4	0.935	0.935	0.944	0.944	0	1.145
TLPP4	1.145	1.145	1.145	1.145	1.145	0

Com esses dados, podemos construir a árvore abaixo, mostrando a proximidade entre ações de mesmo setor e entre setores afins. Você vê como as correlações de preços traduzem a afinidade setorial entre as empresas? Que interessante!

Distâncias subdominantes ultramétricas produzindo
uma hierarquia entre as ações consideradas