



20 e 21 de outubro
Instituto Nacional de Pesquisas Espaciais - INPE
São José dos Campos - SP

Espaços Toblerianos Completos: Recuperando as Visões de Tobler para a Representação dos Espaços de Fluxos na Análise Geográfica

Marcio Azeredo¹, Antônio Miguel Vieira Monteiro²

¹Diretoria de Serviço Geográfico (DSG)
Quartel General do Exército, Bloco "F", 70630-901, Brasília - DF - Brasil.

²Instituto Nacional de Pesquisas Espaciais (INPE)
Av. dos Astronautas 1758, 12227-001, São José dos Campos - SP - Brasil.

{azeredo,miguel}@dpi.inpe.br

Abstract. *Spatial relationships are dynamic, highly anisotropic, and capable of connecting through different types of networks (ex. social, communication and transportation networks). These networks influence the appearance of preferential directions of relationships, providing different representations of space. As a result, requires new methods of exploratory analysis in an attempt to capture this dynamic. This work tries to recover the Bidimensional Regression technique proposed by Tobler in the 70's, together with other techniques of representation, offer a new vision of space study, representing the territory according to network flows. This approach is part of what we call Full Toblerian Spaces (FTS), a proposal for a new mapping in order to provide new opportunities for explanations of the geography of complex socio-spatial phenomena.*

Resumo. *Relacionamentos espaciais são dinâmicos, fortemente anisotrópicos, e passíveis de conexão por intermédio de diferentes tipos de redes (ex. redes sociais, de comunicações e de transportes). Essas redes, por sua vez, influenciam o surgimento de direções preferenciais de relacionamentos, fornecendo diferentes representações do espaço. Como consequência, requer novos métodos exploratórios de análise, na tentativa de capturar essa dinâmica. Este trabalho procura resgatar a técnica da Regressão Bidimensional proposta por Tobler na década de 70, para juntamente com outras técnicas de representação oferecer uma nova visão do espaço em estudo, redesenhando o território em função dos fluxos em suas redes. Essa abordagem é parte do que chamamos de Espaços Toblerianos Completos (ETC), uma proposta para uma nova cartografia com o objetivo de fornecer novas possibilidades de explicações sobre a geografia de fenômenos socioespaciais complexos.*

Palavras-chave: Fluxos, Análise Exploratória, Regressão Bidimensional, Escalonamento Multidimensional.

1. Introdução

Muitos dos conceitos empregados nos atuais Sistemas de Informação Geográfica (SIGs) foram impulsionados pela utilização de uma noção de proximidade baseada em uma métrica de distância euclidiana. Partem do princípio de que os fenômenos espaciais apresentam um comportamento isotrópico, ou seja, que mantêm suas propriedades constantes em qualquer direção dentro do espaço de representação. No entanto, o mundo não é sempre regular e previsível [Tobler 2004]. O conceito de proximidade pode ter diferentes interpretações, havendo a necessidade de novos métodos de análise na tentativa de capturar a complexidade de tais fenômenos.

Quando os SIGs surgiram na década de 60, os dados eram escassos e o custo de processamento elevado [Miller 2000]. Simplificar a forma de representar as relações espaciais era uma solução necessária, conveniente e satisfatória. No entanto, não considerar as propriedades dos objetos espaciais envolvidos, acaba por restringir, em alguns casos, a capacidade de representação e análise de certos tipos de relacionamentos. Fenômenos espaciais são dinâmicos, envolvem relações espaciais fortemente anisotrópicas, e podem ser representados por relações de demanda e atratividade, com origem e destino definidos. Essas ligações configuram os chamados "Espaço de Fluxos"[Gould and White 1986].

Uma importante propriedade dos fluxos é a possibilidade de sua estruturação na forma de diferentes tipos de redes (ex. redes sociais, de comunicações e de transportes). Essas redes servem como dutos na condução dos fluxos [Prager 2008]. Mais do que isso, têm influência na escolha das direções preferenciais de construção dos relacionamentos em função dos custos atribuídos as suas arestas. Isso ocorre, porque a frequência com que as interações surgem varia de acordo com os atributos envolvidos. Normalmente, há uma tendência de que a frequência seja maior onde o custo envolvido for menor, o que possibilita o surgimento de representações distintas do espaço.

O presente estudo parte da premissa de que as transformações geométricas da aplicação da Regressão Bidimensional proposta por Tobler na década de 70, entre outras técnicas de representação, oferecem uma nova visão do espaço em estudo, redesenhando o território em função dos fluxos em suas redes. Com esse propósito, é validada uma metodologia utilizando a Regressão Bidimensional com dados de fluxo da pesquisa de origem e destino [Metro 2008] publicada pela Companhia do Metropolitano de São Paulo - Metrô. O objetivo é oferecer novas possibilidades de explicação sobre a geografia de certos fenômenos.

A abordagem aqui apresentada é uma pequena contribuição para o que chamamos de Espaços Toblerianos Completos (ETC). *Toblerianos* em homenagem a W. R. Tobler, que durante anos dedicou sua pesquisa ao estudo de novas formas de representação de dados espaciais. E *Completos*, por estar aberto a qualquer tipo de métrica, ou método, que contribua para uma cartografia que conduza a novas formas de interpretação das relações entre os lugares e as redes. Entre os fixos que permitem a modificação do espaço, e os fluxos que atravessam, ou ali se instalam [Santos 2006]. No entanto, não há pretensão em esgotar-se o tema, muito pelo contrário.

2. Revisão Bibliográfica

O presente trabalho está inserido na linha de pesquisa de Análise Espacial, e trata sobre técnicas de representação de dados espaciais com implicações na análise exploratória de bases de dados.

Dentre a bibliografia pesquisada para a elaboração do trabalho, verificaram-se algumas iniciativas de estudo sobre a anisotropia dos fenômenos espaciais [Tobler 1979] [Tobler 1993] e sua representação em função de seus atributos. No entanto, a maior parte utilizando a variável "tempo de deslocamento" calculado sobre redes de transporte [Muller 1978] [Spiekermann and Wegener 1994] [Kraak 2009].

Sem sombra de dúvidas, os sistemas de transporte têm grande influência na dinâmica de produção do espaço, e merecem uma atenção especial. Do mesmo modo que as redes onde são estruturados [Prager 2008]. Em Amaral et al. [2006] é possível encontrar uma rica abordagem teórica sobre a importância das redes na produção do espaço.

Matrizes contendo dados de fluxos são utilizadas em alguns trabalhos que abordam a representação de diferentes tipos de fluxos. Fluxos de migração [Tobler 1981] [Tobler 1987], de riquezas e de informação [Walmsley 1883]. Enquanto Tobler dá uma ênfase maior à simbolização, Walmsley preocupa-se com a análise.

O estudo das distorções do espaço produzidas pelas redes de fluxo é realizado neste trabalho com a utilização da Regressão Bidimensional, e de um método de análise multidimensional. A Regressão Bidimensional, inicialmente proposta para comparar configurações espaciais homólogas [Tobler 1994] [Dias 2000], já foi utilizada com esta finalidade em estudos de acessibilidade e mobilidade [Ahmed and Miller 2007] [Cauvin 2002] [Cauvin 2005]. Já o Escalonamento Multidimensional é uma técnica amplamente utilizada na recuperação da configuração espacial embutida em uma matriz de proximidade [Kruskal and Wish 1978] [Borg and Groenen 2005] [Cox and Cox 2000].

3. O *Framework* ETC

3.1. O *framework* ETC e a Terralib

O *framework* ETC foi concebido para ser um repositório de classes desenvolvidas na linguagem C++ funcionando em conjunto com a biblioteca geográfica Terralib [Câmara et al. 2008]. Em um estágio inicial, comportando as classes desenvolvidas para a aplicação da metodologia utilizada neste trabalho, e futuramente agregando outras que funcionem dentro do mesmo conceito (Figura 1).

A biblioteca Terralib foi escolhida como suporte para o desenvolvimento do trabalho, por ser um projeto de software concebido para servir como base para o desenvolvimento colaborativo na comunidade de usuários ou desenvolvedores de SIGs [Casanova et al. 2005]. É um software com código fonte aberto, distribuição livre, e possui seu núcleo de desenvolvimento no Instituto Nacional de Pesquisas Espaciais (INPE), estando disponível para download em: <http://www.terralib.org>.

3.2. Base de dados

A base de dados é composta de duas partes: os dados espaciais e os dados de fluxo. Os dados espaciais devem estar em um formato de armazenamento suportado pela Terralib para

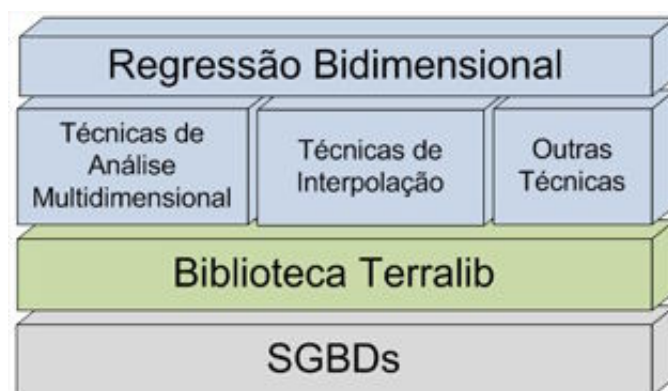


Figura 1. Framework ETC e a Terralib.

que os mesmos possam ser importados para o banco de dados. Aqui, existem três possibilidades: a) os dados são pontos, sendo necessário criar a rede com os relacionamentos presentes na matriz fluxos; b) os dados são linhas, ou seja, já são conhecidos os nós e as arestas da rede; c) os dados são polígonos. Nesse caso, uma alternativa é construção da rede unindo-se cada centróide aos centróides dos polígonos adjacentes. Fato que deve ser considerado na análise dos resultados.

Já os dados de fluxo devem estar estruturados na forma de uma matriz de ordem N (todos se relacionam com todos), contendo os valores que expressam as relações de proximidade (similaridade ou dissimilaridade). Nesse caso, cada matriz terá um máximo de $\frac{N(N-1)}{2}$ relacionamentos, e será analisada individualmente. Para dados de similaridade, quanto maior o valor, maior a noção de proximidade. O raciocínio inverso é aplicado às dissimilaridades.

3.3. Metodologia

O primeiro passo da metodologia (Figura 2) é a construção de uma matriz de dissimilaridades, dada uma origem i , um destino j e o valor numérico das dissimilaridades atribuídas aos relacionamentos (arestas da rede). Caso as relações expressem similaridades, as mesmas devem ser transformadas em dissimilaridades por meio de uma função. Alguns exemplos de funções podem ser encontrados em [Cox and Cox 2000].

Uma vez construída a matriz de dissimilaridades, o Escalonamento Multidimensional (MDS) é responsável por recuperar a representação espacial embutida em uma matriz de dissimilaridades [Friedman and Kohler 2003]. Para os propósitos atuais do trabalho, foi utilizado o Escalonamento Multidimensional Clássico (CMDS), que associa os próprios valores de dissimilaridade às distâncias originais dos objetos no espaço.

Em Hair et. al. [1998] são definidos alguns métodos de determinação da melhor dimensão de representação da solução do CMDS. Neste trabalho, existe uma imposição de que a solução tenha duas dimensões, visando favorecer a análise com os dados originais.

A aplicação da Regressão Bidimensional parte de duas configurações espaciais homólogas situadas em dois sistemas coordenados distintos. O sistema da coordenadas do espaço de referência (Z) e o sistema da coordenadas da solução do CMDS (W).

Segundo Cauvin [2005], a Regressão Bidimensional pode ser dividida em duas fa-

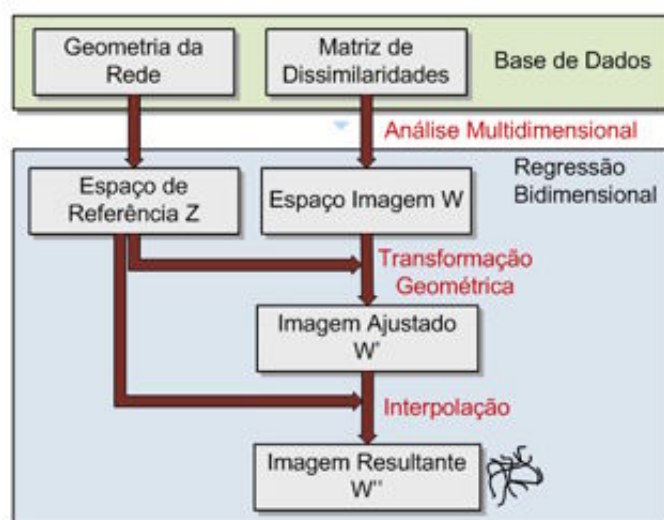


Figura 2. Resumo da metodologia. Adaptado de: [Cauvin 2005].

ses principais: um ajustamento das observações e uma interpolação de forma a generalizar os resultados para todo o espaço.

A primeira fase consiste da aplicação de uma transformação geométrica linear seguida de um ajustamento das observações pelo Método dos Mínimos Quadrados (MMQ). As transformações lineares mais utilizadas são: a transformação de similaridade, onde são considerados quatro parâmetros (dois de translação, um de rotação e um de escala) e a transformação afim geral, onde são considerados seis parâmetros (dois de translação, dois de escala, um de rotação e um de cisalhamento). Neste trabalho foi utilizada a transformação afim, por esta apresentar os melhores resultados. Desse modo, as coordenadas de W' permitem construir os vetores principais de distorção que serão utilizados na próxima etapa.

A segunda fase procura extrapolar os vetores principais de distorção a todos os pontos pertencentes ao espaço. A interpolação é feita utilizando o método de interpolação *Thin-Plate Splines* (TPS), que é um método que procura minimizar a intensidade das distorções envolvidas na transformação entre as duas configurações de pontos. Desse modo, as distorções são expressas pela Regressão Bidimensional.

Cabe ressaltar, que as escolhas do método de análise multidimensional e de interpolação foram opções dos autores. Nada impede que outros métodos sejam analisados e agregados futuramente.

4. Estudo de caso: a pesquisa O/D 2007

A Pesquisa Origem e Destino, ou simplesmente Pesquisa O/D, é realizada desde 1967 na Região Metropolitana de São Paulo (RMSP), com periodicidade de dez anos, e tem por objetivo o levantamento de informações atualizadas sobre as viagens realizadas pela população da metrópole em dia útil típico [Metro 2008]. Os resultados da pesquisa estão condensados em um conjunto de planilhas com dados de origem, destino, modo principal de transporte (ex. trem, ônibus, metrô), tipo de transporte (ex. coletivo, individual), quantidade de viagens e o tempo médio das mesmas, entre as 460 zonas de pesquisa.

A divisão em 460 zonas foi uma decisão operacional do Metrô. No entanto, trata-se de um volume de dados considerável para ser utilizado de uma única vez. Sendo assim, para o estudo de caso optou-se pela seleção de uma amostra de 10 zonas localizados na região central da RMSP (Figura 3). São elas: Sé, Parque Dom Pedro, Praça João Mendes, Ladeira da Memória, República, Santa Efigênia, Luz, Bom Retiro, Ponte Pequena e Canindé.



Figura 3. Divisão da RMSP em áreas de pesquisa.

Os experimentos conduzidos neste estudo foram realizados considerando os seguintes atributos da amostra: 1) tempo médio mínimo de duração das viagens. 2) número total de viagens. Os pontos de referência para a construção da rede foram calculados, em ambos os experimentos, a partir dos centróides das zonas.

4.1. Experimento 1

O tempo médio mínimo utilizado neste experimento corresponde ao tempo médio de viagem mais curto entre dois pontos da amostra. Todos os meios de transporte foram considerados para este fim. Como a variável tempo representa uma dissimilaridade, não há a necessidade de utilização de nenhuma transformação a priori, sendo os dados da matriz utilizados diretamente pelo CMDS.

Para a construção da rede de relacionamentos foram interligados os centróides dos polígonos adjacentes.

O passo seguinte é a montagem da matriz superior de conexão com os valores de tempo tomados aos pares adjacentes (Figure 4). Um algoritmo de caminho mínimo determina o menor tempo de duração (caminho mínimo) das viagens entre todos os pontos da amostra.

Após a aplicação da metodologia, foi utilizada uma malha para representar a distorção do espaço em função do atributo tempo (Figura 5).

A figura 6 mostra o espaço resultante como efeito das distorções produzidas nos polígonos da amostra.

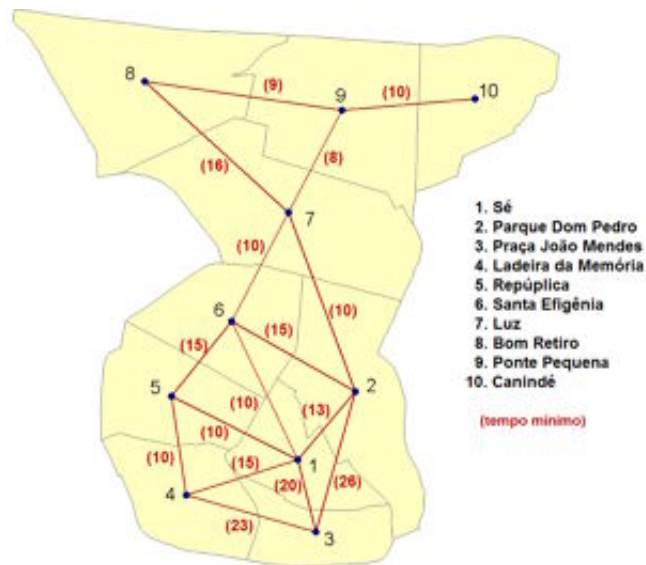


Figura 4. Rede adjacente valorada pelo tempo médio mínimo (em minutos).

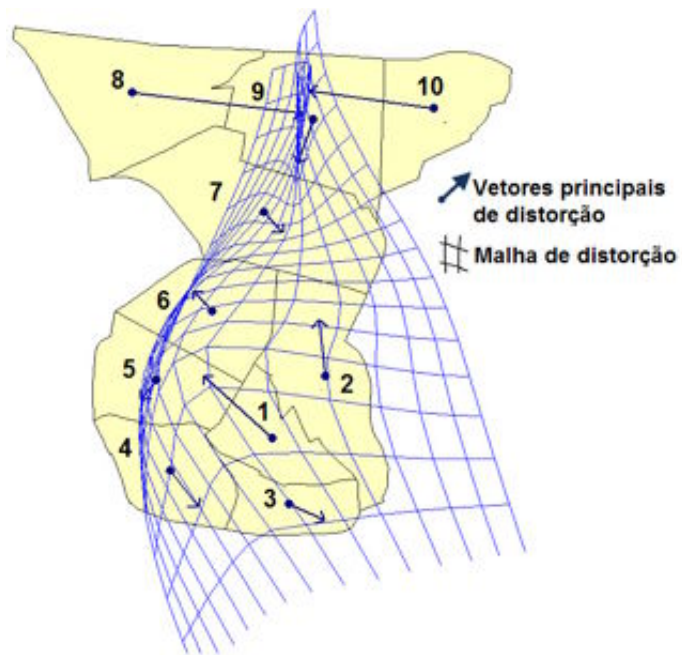


Figura 5. Malha de distorção do experimento 1.

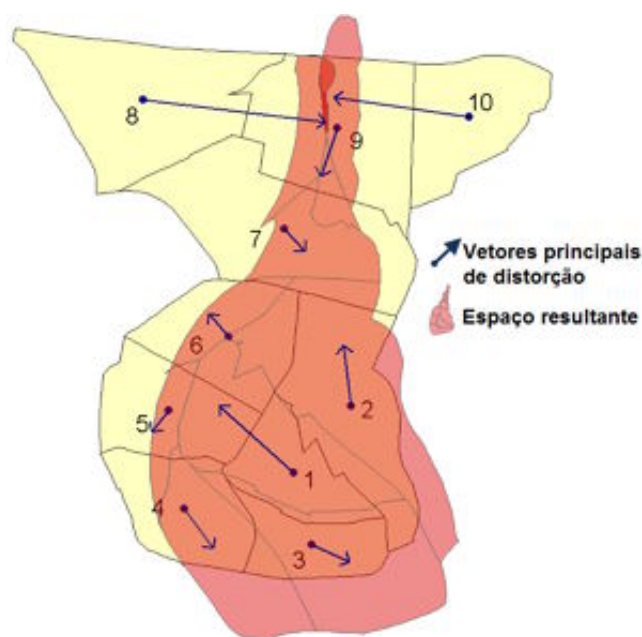


Figura 6. Espaço resultante do experimento 1.

Analisando as figuras 4, 5 e 6, é possível fazer algumas considerações. A contração do espaço observada na região superior da figura 5 e 6 é compatível com os baixos tempos de deslocamento na região. Principalmente entre os pares 8-9, 7-9 e 9-10. O ponto 9 pode ser considerado um ponto de acesso rápido na região. Já na parte inferior, são observados os maiores tempos de acesso, principalmente nos pares 1-4, 2-3 e 3-4. Isso faz com que ocorra uma distensão ao longo da parte mais externa (2 e 3). Em contrapartida, as ligações 1-5, 1-6 e 4-5 são mais rápidas e promovem a contração das áreas 4, 5 e 6. Pode-se dizer, de uma maneira geral, que as áreas de contração do espaço indicam onde é possível chegar mais rápido. Os resultados não possibilitam afirmar se a infraestrutura comporta a demanda pelo serviço. No entanto, como o sistema é mais rápido naquela região, existe a possibilidade de que com o tempo ocorra algum tipo de sobrecarga.

4.2. Experimento 2

O número total de viagens considerado neste experimento corresponde à soma de todos os meios de transporte pesquisados. Os valores foram condensados de forma a considerar a demanda total pelo serviço.

Neste caso, a variável número de viagens representa similaridade, havendo a necessidade da utilização de uma transformação a priori. Com esse propósito, foi utilizada uma transformação simples que inverteu a escala de valores.

O número total de viagens considerado neste experimento corresponde à soma de todos os meios de transporte pesquisados. Os valores foram condensados de forma a considerar a demanda total pelo serviço.

Uma primeira vista da Figura 7, já possibilita verificar que se adensamento da rede é bem maior que o experimento 1.

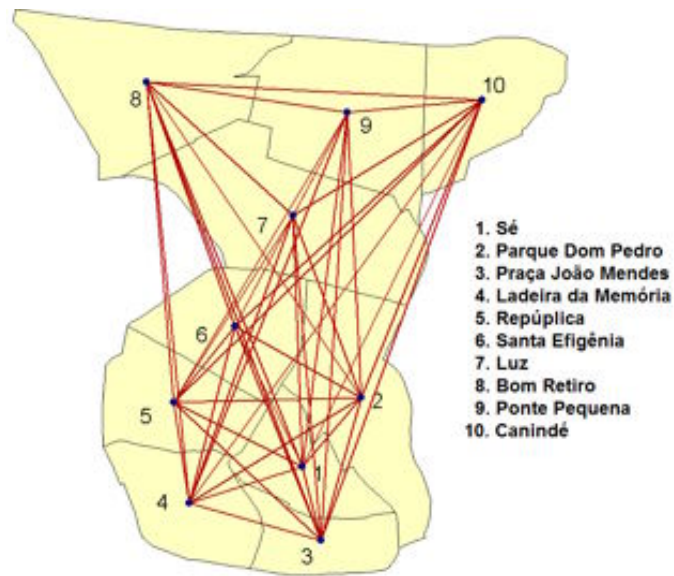


Figura 7. Rede valorada número total de viagens.

A malha de distorção do espaço em função do número total de viagens pode ser observada na figura 8.

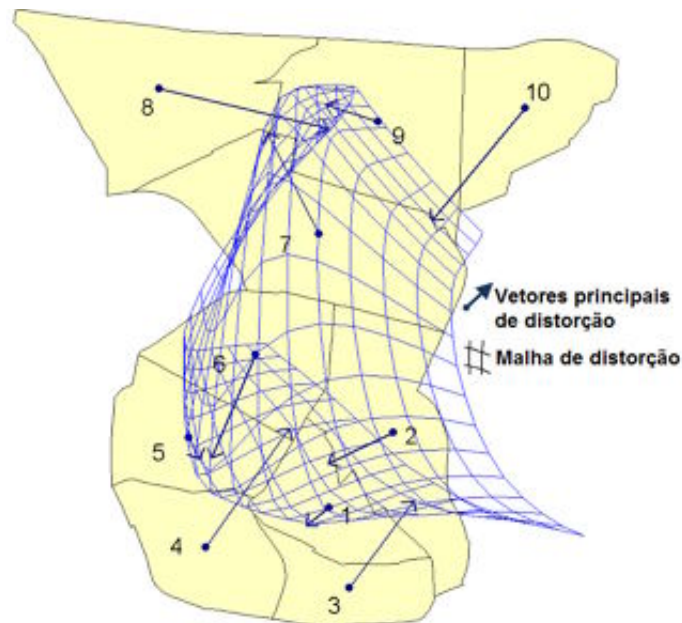


Figura 8. Malha de distorção do experimento 2.

A figura 9 mostra o espaço resultante como efeito das distorções produzidas nos polígonos da amostra. É possível observar que as magnitudes das distorções é maior no experimento 2, possivelmente por lidar com uma faixa de valores de atributos mais extensa.

Confrontando os resultados com os dados da matriz, verifica-se que existe uma grande demanda por viagens nas áreas compreendidas pelos relacionamentos 1-2, 1-3, 1-6, 5-6 e 7-8, 7-9, indicando que essas áreas necessitam de um cuidado no planejamento

de sua infraestrutura de transporte, para que no futuro tenham a capacidade de suportar a demanda solicitada

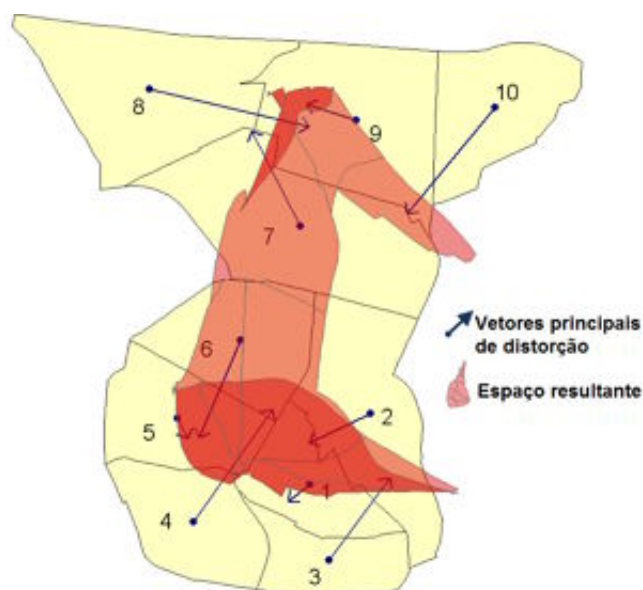


Figura 9. Espaço resultante do experimento 2.

5. Conclusões e Perspectivas

Este trabalho trata da utilização de novos métodos de análise exploratória capazes de oferecer novas possibilidades de representar e visualizar dados espaciais. Mesmo que os experimentos tenham sido realizados, em uma abordagem inicial, com algum grau de simplificação, acreditamos que as idéias são úteis para aquilo que se propuseram: capturar as distorções produzidas pelos fluxos em suas redes.

O *framework* FTS atualmente é restrito à aplicação da metodologia descrita no trabalho. No entanto, existem muitas possibilidades de ampliação de sua capacidade de representação e análise. Alguns dos exemplos são: implementação de novas técnicas de análise multidimensional e de interpolação; pesquisa de alternativas para a utilização de grandes bases de dados de fluxo; tratamento da ausência de relacionamentos nas matrizes de proximidades; implementação do framework como um serviço de processamento na web (WPS); entre outras.

O grande desafio que enfrentamos é a transição dos mapas que apresentam respostas, para os mapas que fomentam a busca pelas mesmas [MacEachren and Monmonier 1992].

5.1. Agradecimentos

Os autores deste trabalho agradecem à Companhia do metropolitano de São Paulo (Metrô) pela disponibilização dos dados referentes à Pesquisa O/D 2007. É um grande estímulo poder colaborar com a parceria entre o Metrô e o INPE (Programa Espaço e Sociedade), na busca de novas possibilidades analíticas de uso dos dados da pesquisa OD do Metrô.

Referências

- Ahmed, N. and Miller, H. J. (2007). Time-space transformations of geographic space for exploring, analysing and visualizing transportation systems. *Journal of Transport Geography*, (206):2–17.
- Borg, I. and Groenen, P. J. F. (2005). *Modern Multidimensional Scaling*. Springer, Estados Unidos da América.
- Casanova, M. A., Câmara, G., Davis Jr, C. A., Vinhas, L., and Queiroz, G. (2005). *Bancos de Dados Geográficos*. MundoGeo, Curitiba.
- Cauvin, C. (2002). Cognitive and cartographic representations. *Cybergeo*, (206).
- Cauvin, C. (2005). A systemic approach to transport accessibility. *Cartography and Geographic Information Science*, (311).
- Câmara, G., Vinhas, L., Queiroz, G., Ferreira, K., Monteiro, A. M. V., Carvalho, M. T., Cartaxo, R., Casanova, M. A., and Freitas, U. M. (2008). Terralib. In Hall, G. B. and Leahy, M., editors, *Open Source Approaches to Spatial Data Handling*. Springer-Verlag, Berlin.
- Cox, F. C. and Cox, M. A. A. (2000). *Multidimensional Scaling*. Chapman and Hall, Boca Raton.
- Dias, M. H.; Cauvin, C. A. M. F. (2000). Comparação de configurações cartográficas através da regressão bidimensional. *Finisterra*, 35(69):95–107.
- Friedman, A. and Kohler, B. (2003). Bidimensional regression. *Psychological Methods*, 8(4):468–491.
- Gould, P. and White, R. (1986). *Mental Maps*. Routledge, USA.
- Kraak, M. and Maceachren, A. M. (2009). The shrivelled usa. *Journal of Transport Geography*, 17(6):433–439.
- Kruskal, J. B. and Wish, M. (1978). *Multidimensional Scaling*. Eric M. Uslaner.
- MacEachren, A. and Monmonier, M. (1992). Introduction. *Cartography and Geographic Information Systems*, 19(4):197–200.
- Metro (2008). Pesquisa origem e destino 2007: Região metropolitana de São Paulo.
- Miller, H. J. (2000). Geographic representation in spatial analysis. *Journal of geographical systems*, 2(1):55–60.
- Muller, J. C. (1978). The mapping travel time in Edmonton. *Canadian Geographer*, 22(3):195–210.
- Prager, S. D. (2008). Complex networks for representation and analysis of dynamic geographies. In Hornsby, K. S. and Yuan, M., editors, *Understanding Dynamics of Geographic Domains*. CRC Press, Estados Unidos da América.
- Santos, M. (2006). *A natureza do Espaço*. EdUSP, São Paulo.
- Spiekermann, K. and Wegener, M. (1994). The shrinking continent. *Planning and Design*, 21(6):653–673.

- Tobler, W. R. (1979). A transformational view of cartography. *The American Cartographer*, 6(2):101–106.
- Tobler, W. R. (1981). A model of geographical movement. *Geographical Analysis*, 13(1):1–20.
- Tobler, W. R. (1987). Experiments in migration mapping by computer. *The American Cartographer*, 14(2):155–163.
- Tobler, W. R. (1993). *Three Presentations on Geographical Analysis and Modeling*. National Center for Geographic Information and Analysis, University of California.
- Tobler, W. R. (1994). Bidimensional regression. *Geographical Analysis*, 26(3):187–212.
- Tobler, W. R. (2004). On the first law of geography. *Annals of the Association of American Geographers*, 94(2):304–310.
- Walmsley, D. J. (1883). Public information flows in rural australia. *Environment and Planning A*, 15(2):255–263.