

# Análise exploratória da pegada digital dos turistas para a identificação de padrões espaciais e temporais em destinos urbanos\*

Luis Encalada\*\*  
Jorge Rocha\*\*\*  
Carlos Ferreira\*\*\*\*

*Recebido 08 de maio de 2018; aceite 26 de setembro de 2018*

## Resumo

No presente trabalho, utilizaram-se as fotos georreferenciadas dos visitantes da cidade de Lisboa, partilhadas na rede ‘Panoramio’ entre 2007 e 2014, como proxy para analisar a distribuição espacial e temporal dos mesmos dentro da cidade. O conjunto total de dados (> 75.000 fotografias) foi segmentado em visitantes e locais, com base nas marcas temporais, resultando numa amostra de 17.604 fotos tiradas por > 5.000 utilizadores. A evidência empírica sugere que a distribuição espacial dos visitantes não é homogénea. Complementarmente, avaliamos a relação espacial entre o padrão observado (a aglomeração geográfica de visitantes) e um conjunto de 24 variáveis associadas à oferta turística da cidade. Através da análise de regressão linear múltipla, verificou-se que são os ‘Monumentos de interesse público’, os elementos que apresentam maior atratividade e, apurou-se, considerando os fatores explicativos selecionados, que existem áreas do centro histórico, cujo potencial turístico está subestimado.

\* Este artigo é apresentado a partir dos resultados da Dissertação de Mestrado, intitulada: “‘Big Data’ e Redes Sociais na Análise Espacial de Padrões Turísticos em Contexto Urbano. O caso de Lisboa”, defendida em 26 de setembro de 2014.

\*\* Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, Rua Branca Edmée Marques, 1600-276, Lisboa, Portugal, e-mail: [luisencalada@campus.ul.pt](mailto:luisencalada@campus.ul.pt)

\*\*\* Instituto de Geografia e Ordenamento do Território, Universidade de Lisboa, e-mail: [jorge.rocha@campus.ul.pt](mailto:jorge.rocha@campus.ul.pt)

\*\*\*\* Departamento de Geografia e Turismo, Faculdade de Letras, Universidade de Coimbra; Colégio de S. Jerónimo, 3004-530 Coimbra, Portugal, e-mail: [carlos.ferreira@uc.pt](mailto:carlos.ferreira@uc.pt)

Palavras-chave: *Geotagged fotos, Turismo urbano, Redes sociais, Big Data, Análise espaciotemporal.*

## Resumen

En este artículo se utilizaron las fotos georreferenciadas, compartidas por usuarios de la red “Panoramio” entre 2007 y 2014, como un *proxy* para analizar la distribución espacial y temporal de los visitantes, en la ciudad de Lisboa. El conjunto total de datos (> 75.000 fotografías) fue segmentado en turistas y locales, con base en las marcas temporales, resultando una muestra de 17.604 fotos tomadas por > 5.000 usuarios. La evidencia empírica sugiere que la distribución espacial de los visitantes no es homogénea. Además, exploramos la relación espacial entre el patrón observado (la aglomeración geográfica de visitantes) y un conjunto de 24 variables asociadas a la oferta turística de la ciudad. A través del análisis de regresión lineal múltiple, se verificó que son los “monumentos de interés público”, los elementos que presentan mayor atractivo, considerando los factores explicativos seleccionados, y que existen áreas del centro histórico cuyo potencial turístico está subestimado.

Palabras clave: *Fotos geotiquetadas, Turismo urbano, Redes sociales, Big Data, Análisis Espacio-temporal.*

## Abstract

In this paper, we use online digital footprints (i.e., geotagged photos from ‘Panoramio’) of Lisbon’s visitors, between 2007 and 2014, as a sensitive proxy for exploring tourist spatiotemporal behavior within the city. We identify the visitors’ photos from the overall dataset (> 75,000 photos) by looking at photos’ timestamps, resulting in a sample of 17,604 photos from > 5,000 users. This empirical evidence suggests that the spatial distribution of visitors is not homogenous. Then, we assess the spatial relation between the observed pattern (geographical agglomeration of visitors) and an initial set of 24 variables regarding city tourism offer. Findings from the multiple regression analysis, show that ‘Monuments of public interest’ are the elements with higher attractiveness. Furthermore, some areas located within the historic center appear underestimated, despite their significant tourism appeal considering the selected variables.

Key words: *Geotagged photos, Urban Tourism, Social networks, Big data, Spatiotemporal analysis.*

## Introdução

O entendimento de como os turistas usam o espaço nos ambientes urbanos tem aplicações práticas na administração e planejamento do turismo nos locais de destino

(Lau e McKercher, 2006), podendo aproveitar-se como contributo de suporte para a organização dos serviços essenciais ao funcionamento da atividade, a criação de novos produtos, a gestão de impactos nas suas diversas formas, entre outros fins (Hayllar *et al.*, 2008).

Segundo Ashworsth e Page (2011), apesar da importância quantitativa deste tipo de turismo, é curioso a pouca atenção que tem sido dada ao entendimento de como os turistas realmente usam as cidades. Tal como referido por vários autores (Kádár, 2014; Pearce, 2001), o comportamento espacial da procura (no contexto urbano) ainda não é totalmente compreendido devido à falta de dados com este nível (local) de discriminação, justificada pelas dificuldades na sua quantificação, *i.e.*, limitações na aplicação de métodos científicos para a monitorização de turistas.

A par do desenvolvimento das novas tecnologias de informação e de comunicação, têm vindo a surgir novas fontes de dados, menos tradicionais, que podem ser aproveitadas para o estudo do turismo urbano, particularmente, para a análise do comportamento espacial e temporal dos turistas. Os serviços online de partilha de fotos (*e.g.*, Flickr, Instagram), de mensagens instantâneas (*e.g.*, Twitter), de avaliação de consumidores (*e.g.*, Tripadvisor) e outras plataformas envolvidas num ambiente colaborativo de partilha de conteúdos, formam parte integrante destas novas fontes, que se caracterizam pelos altos níveis de participação e pela quantidade de informação gerada. Estas plataformas expandem o cenário da quantidade/diversidade de fontes disponíveis, e deixam de lado a condição em que a informação era exclusivamente produzida pelas autoridades oficiais e através de métodos científicos clássicos. Estas fontes não oficiais, enquadradas no campo das *social media*, têm promovido os próprios indivíduos como geradores de informação.

Os turistas, quando utilizam as redes sociais, deixam grandes quantidades de rastros digitais, resultando num conjunto multidimensional de dados conhecido como “Big Data” (Buhalis e Amaranggana, 2014). Assim, como referido por Buhalis e Law (2008) e Munar e Jacobsen (2014), denota-se uma tendência crescente no número de turistas que partilham as suas experiências através da publicação de recomendações, opiniões, fotos ou vídeos sobre um destino, atividade ou serviço, em sítios de redes sociais.

Recentemente, a informação disponível nas redes sociais e noutras plataformas digitais tem vindo a ser aproveitada como complemento aos dados provenientes de estatísticas oficiais, nomeadamente, por serem de livre acesso e com um nível elevado de resolução espacial e temporal (Goodchild e Li, 2012). Estes dados, originados a partir da comunidade, proporcionam novas oportunidades para as ciências geográficas e para o estudo de fenómenos sociais, podendo ser utilizados como *proxy* para medir a capacidade de atracção dos locais e para analisar a distribuição espacial dos turistas num determinado local de destino.

O desenvolvimento deste artigo prende-se, por um lado, na análise do comportamento espacial e temporal dos visitantes da cidade de Lisboa através da aplicação de métodos de estatística espacial, com recurso a dados extraídos da rede social ‘Panoramio’ e, por outro lado, com a exploração das relações espaciais que podem explicar o padrão observado nos turistas.

## Dados e métodos

### *Lisboa: A região e a cidade*

De acordo com as Estatísticas do Turismo do Instituto Nacional de Estatística (INE, 2015), para o ano 2014, a Área Metropolitana de Lisboa (AML) concentrou 24,95% do total das dormidas no país. A região está entre as primeiras escolhas dos principais mercados externos, sendo o destino preferido pelos espanhóis, brasileiros, norte-americanos e italianos. Lisboa é a cidade do país que apresenta o maior número de dormidas e de hóspedes, com mais de 9 e de 3,5 milhões, respetivamente (INE, 2015). A estada média dos hóspedes estrangeiros na cidade é de 2,5 noites, abaixo da média da AML (2,6) e de Portugal Continental (3,1 noites).

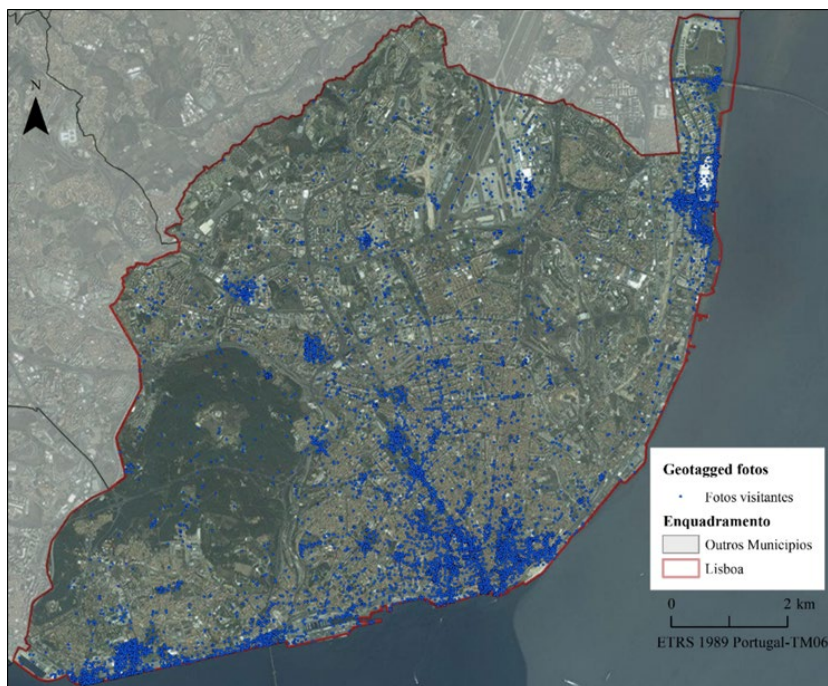
### *Geotagged fotos de redes sociais como fonte de dados*

Durante a sua visita, os turistas, em função das suas motivações, vão consumir as experiências oferecidas pela cidade. Ainda que não se verifique um consumo ativo, haverá pelo menos um consumo indiciário, através do ato de tirar fotografias (Kádár, 2014). Assim como referido por Urry e Larsen (2011), a maioria dos turistas tiram fotografias com o objetivo de criar registos da sua visita. As fotografias são agora amplamente produzidas, consumidas e partilhadas via internet, especialmente através das redes sociais (Lo *et al.*, 2011).

No contexto urbano, vários estudos têm utilizado *geotagged* fotos provenientes de redes sociais para analisar o comportamento espacial dos turistas e identificar os locais de concentração (Encalada *et al.*, 2017; García-Palomares *et al.*, 2015; Tammet *et al.*, 2013), assim como os padrões de mobilidade (Jin *et al.*, 2017). Recorrendo à informação publicada e disponibilizada na rede de partilha de fotos ‘Panoramio’, foram retiradas 75.112 fotografias localizadas no concelho de Lisboa, que correspondem as imagens carregadas pelos utilizadores (da rede) entre os anos 2007 e 2014.

O método para diferenciar as fotos tiradas por visitantes e locais assenta nos trabalhos realizados por Girardin *et al.* (2008), Kádár (2014), García-Palomares *et al.* (2015), Encalada *et al.* (2017). Os autores calcularam a diferença de dias entre a marca temporal (*timesteps*) da primeira e última foto carregadas pelos utilizadores, dentro da área de estudo. Se o número de dias entre a primeira e última foto excedesse a média do tempo de visita do destino (neste caso, < 4 dias), as fotos eram

consideradas como pertencentes a utilizadores locais. Após o processo de filtragem, apuraram-se um total de 17.604 fotos tiradas por cerca de 5.000 utilizadores considerados como visitantes (Figura 1).



**Figura 1.** Fotos dos turistas de Lisboa, entre os anos 2007 e 2014.

### ***Análise dos padrões turísticos em Lisboa***

Na Figura 2, apresenta-se o esquema processual da aplicação dos métodos estatísticos, sintetizando a sequência dos procedimentos e os propósitos da utilização dos métodos.

O ponto de partida é validar a autocorrelação espacial e determinar se a distribuição dos dados é dispersa, concentrada ou aleatória. Quando avaliada a autocorrelação espacial, surge o interesse em perceber a evolução do grau de aglomeração (*clustering*), em função da variação da distância entre observações, e em comparar os padrões em diferentes períodos, com o fim de avaliar quais as suas mudanças no tempo. Apesar de, visualmente, se notar uma tendência para os visitantes se concentrarem nas áreas turísticas, procura-se, com a aplicação destes primeiros métodos, verificar se o padrão é estatisticamente significativo e, como tal, suportado por um processo espacial subjacente.



**Figura 2.** Esquema processual da aplicação dos métodos de análise espacial e temporal.

Sendo estas métricas derivadas das características globais da amostra, constatando-se a tendência comum das observações em se concentrarem, o que resta é identificar quais os locais de concentração (mais uma vez, considerando a sua significância estatística). Para o efeito, deve-se recorrer a métodos que consigam interpretar a heterogeneidade da autocorrelação espacial.

Assim sendo, optou-se por uma análise de *clusters* e *outliers*, baseada no *Local Moran Index* (Anselin, 1995), para determinar quais as aglomerações estatisticamente significativas, sejam compostas por valores semelhantes elevados (*High-High clusters*) ou baixos (*Low-Low clusters*), e os *outliers* espaciais. O cálculo do índice implica estabelecer a noção de proximidade entre observações, sendo alcançada, neste caso, através da criação de uma matriz de pesos espaciais.

Todavia, importa discorrer qual a influência da componente temporal na conceção dos *clusters* espaciais. É neste sentido que, através da ferramenta *Emerging hot spot analysis* (ESRI, 2016), se analisa, para além da relação espacial

entre as fotografias, a sua indexação temporal face ao período em que foram carregadas. Assim, em complemento à localização dos focos, obtém-se como resultado o seu padrão sazonal (*e.g.*, se os locais de concentração são consecutivos, esporádicos, etc.).

Previamente à identificação dos padrões espaciais, dedica-se uma parte da análise, à descrição das características da distribuição espacial dos visitantes, com recurso a duas métricas, o centro médio e a elipse do desvio padrão. Por intermédio destas, tenta-se representar a compacidade e a orientação espacial do conjunto de dados. Com o intuito de avaliar se as características da distribuição diferem de um ano para outro, as medidas foram aplicadas para cada ano, de forma segmentada.

Deve-se mencionar que a análise exploratória de dados foi operacionalizada através de várias ferramentas disponíveis no *software* ArcMap. Os parâmetros utilizados para cada uma das ferramentas são mencionados com detalhe no que resta desta secção.

Por fim, exploram-se as relações espaciais existentes entre o padrão observado dos visitantes e um conjunto de 24 variáveis selecionadas que representam parte da oferta turística da cidade (*e.g.*, Hotelaria, arquitetura religiosa, miradouros, etc.). O que se pretende com a aplicação da análise de regressão linear múltipla, é evidenciar se as áreas com potencial turístico são também as que concentram mais visitantes, *i.e.*, se são as mais cativantes para os visitantes.

#### *Centro médio e elipse do desvio padrão*

O cálculo do centro médio e da elipse do desvio padrão é simples, tanto em termos computacionais como na sua parametrização. No caso da elipse do desvio padrão, é preciso definir os desvios padrões a considerar para o seu cálculo. Neste caso, as elipses foram apuradas considerando um desvio padrão. Para esta análise, recorreu-se ao posicionamento original das fotografias sem serem agregadas.

#### *Índice do vizinho mais próximo*

Para o cálculo do índice é necessário definir a extensão da área de estudo (uma vez que é utilizada para medir a distância média observada e esperada), e o critério de proximidade. A extensão corresponde à área da cidade  $\approx 86 \text{ km}^2$ . No que diz respeito ao critério de proximidade, a distância entre os vizinhos mais próximos será determinada a partir do segmento de recta que os une, a distância Euclidiana. Para esta análise, utilizou-se igualmente a localização original das fotografias sem serem agregadas.

#### *K-function (Multi-distance spatial cluster analysis)*

Os dados a utilizar nesta análise correspondem às fotografias de dois dos oito anos disponíveis (2007 e 2012). Estes anos foram escolhidos por apresentarem uma

quantidade de fotos superior à da média anual. O número de intervalos (de distância), para os quais a análise foi efectuada é de 20. Quanto maior o número, mais suave será a curva dos valores observados da função  $L(d)$ , permitindo identificar com mais facilidade a distância à qual o grau de *clustering* é maior.

Na distância de partida estabeleceu-se a distância mínima entre as observações e o primeiro vizinho mais próximo, para assegurar que, no início da análise, todas as observações tenham pelo menos um vizinho, garantindo assim a significância dos resultados. A distância resultante foi de 50 m.

O valor do incremento para os intervalos foi estabelecido, tomando como referência a distância média resultante da análise do índice do vizinho mais próximo, *i.e.*, a distância média entre as observações e os seus vizinhos mais próximos (65 m). No que concerne aos limites de confiança -altos e baixos- estes foram determinados com base num teste de 999 permutações, obtendo assim, um nível de confiança de 99,9%. Para esta análise, fez-se novamente uso da localização original das fotografias não-agregadas.

#### *Análise de clusters e outliers (ACO)*

As observações foram previamente agregadas em unidades hexagonais de aproximadamente 2.350 m<sup>2</sup>. O valor resultante da célula corresponde à soma do número de fotografias localizadas dentro da própria área. A fim de executar a análise, torna-se necessário definir: i) a área circundante (área que abrange as entidades vizinhas a comparar) a cada *target feature*; e ii) a natureza da relação espacial entre as entidades. A relação espacial foi operacionalizada através de uma matriz de pesos espaciais, sendo baseada no declínio da função da distância (euclidiana). Assim, a influência das *features* vizinhas vai decrescendo, na medida em que a distância entre elas (*target feature* e  $n$  vizinhas) vai aumentando. Assumindo que nem todas as observações dentro da área de estudo estão relacionadas espacialmente, estabeleceu-se um limite de distância para o raio de influência da vizinhança, igual a 135 m (toma-se como referência o centroide da célula).

#### *Análise de pontos quentes emergentes (APQE)*

Para executar a APQE, deve-se primeiro agregar os dados de entrada num cubo espaciotemporal (em formato NetCDF). Optou-se por uma agregação de um mês a fim de mitigar, ou mesmo omitir, o desvio/erro potencialmente presente na referência temporal das fotografias. A dimensão espacial das caixas de agregação corresponde a mesma área das unidades utilizadas para a ACO. As fotografias foram agregadas em 65.192 caixas espaciotemporais, cobrindo a área de estudo e um período de 91 meses, de Janeiro 2007 a Julho 2014.



O método APQE tem como base a estatística de *Getis-Ord Gi\** (Getis e Ord, 1992). Para cada caixa de agregação do cubo, a ferramenta calcula o índice, considerando o valor da própria caixa (*target*) e das caixas na sua vizinhança (estabelecida através dos parâmetros de distância espacial e temporal). Assim, uma caixa de agregação com valor elevado será um ponto quente estatisticamente significativo, a menos que as suas vizinhas no espaço-tempo também apresentem valores elevados. Para estabelecer o limite da vizinhança, considerou-se a mesma distância da ACO, de 135 m da caixa *target* (toma-se como referencia o centroide da caixa, num contexto 2D). A vizinhança temporal foi estabelecida para considerar 12 meses para a mesma localização (caixa).

### *Análise de regressão lineal múltipla*

Na tentativa de explorar as relações espaciais existentes entre o padrão observado dos visitantes e um conjunto de variáveis que representam parte da oferta turística da cidade (e.g., Hotelaria, arquitetura religiosa, miradouros, etc.), aplicou-se uma análise de regressão linear múltipla, podendo assim, evidenciar se as áreas com potencial turístico (ou melhor, áreas mais próximas aos elementos da oferta) são também as que concentram mais fotos.

Um dos primeiros aspetos a ser considerados na esquematização de um modelo de regressão linear é a seleção dos fatores explicativos (variáveis) a utilizar, dado que uma variável independente adicional pode melhorar a previsão da variável dependente.

De acordo com a informação disponibilizada pelo sítio internet da Câmara Municipal de Lisboa referente à atividade turística, foram selecionadas 24 variáveis (Tabela 1) para serem utilizadas como fatores explicativos da aglomeração geográfica dos visitantes. Todas as variáveis (*i.e.*, elementos da oferta representados pela sua localização espacial) foram transformadas em escalas de razão, através da aplicação do inverso da distância euclidiana ao elemento analisado mais próximo. Como a unidade de medida é sempre a mesma (distância em metros) não se revelou necessário proceder à normalização dos valores.

A melhoria do desempenho preditivo de um modelo de regressão linear não está só relacionada com a correlação de uma determinada variável com a variável dependente, mas também com a correlação dessa variável com as demais variáveis independentes do modelo. Deve-se, pois, avaliar as possíveis dependências entre os regressores. Em situações onde essas dependências forem fortes, pode-se afirmar que existe multicolinearidade.

A presença de multicolinearidade pode ser detetada de várias maneiras. Uma técnica muito usada é o valor de tolerância ou o seu inverso, designado por fator de inflação de variação (*Variance Inflation Factor* [VIF]), cujos valores elevados

indicam a existência de multicolinearidade. O VIF é uma medida do grau em que cada variável independente é explicada pelas demais variáveis independentes. Quanto maior for o fator de inflação da variância, mais severa será a multicolinearidade. Como regra geral, as variáveis explicativas com valores de VIF superiores a 7,5 devem ser removidas (uma a uma) do modelo de regressão. Das 24 variáveis pré-selecionadas, 12 foram excluídas por apresentarem valores de VIF > 7,5 (Tabela 1).

**Tabela 1**  
**Variáveis explicativas pré-selecionadas, e os respectivos valores de VIF**

Código	Variável	VIF	Código	Variável	VIF
V1	Arquitetura Civil	5,39	V13	Hotelaria (Alojamentos)	6,15
V2	Arquitetura Industrial	5,87	V14	Igrejas	16,89
V3	Arquitetura Militar	20,85	V15	Mercados	10,3
V4	Arquitetura Nobre	16,82	V16	Miradouros	58,23
V5	Arquitetura Premiada	4,5	V17	Monumentos de Interesse Público	3,96
V6	Arquitetura Religiosa	45,07	V18	Monumentos Nacionais	58,77
V7	Cemitérios	3,49	V19	Museus	6,24
V8	Conjuntos de Interesse Público	104,9	V20	Parques Infantis	4,13
V9	Docas de Recreio e Marinas	6,7	V21	Parques Merendas	359
V10	Elevadores e Ascensores	114,6	V22	Parques Recreativos	442,8
V11	Feiras	22,37	V23	Parques Urbanos	2,99
V12	Geomonumentos	5,69	V24	Teatros	4,27

**Fonte:** Elaboração própria.

Posto isto, foi realizada uma análise exploratória em que foram corridos modelos que permitissem a existência de 1 a 5 variáveis independentes. Foram testadas 35.526 combinações das quais 6.866 apresentaram um  $R^2$  ajustado superior a 0,9. Dentro deste grupo, mas restrito, os 38 modelos com melhor desempenho recorreram sempre a 5 variáveis, com desempenhos preditivos entre 0,95 e 0,96. Cruzando as variáveis que foram mais vezes selecionadas por esses 38 modelos com

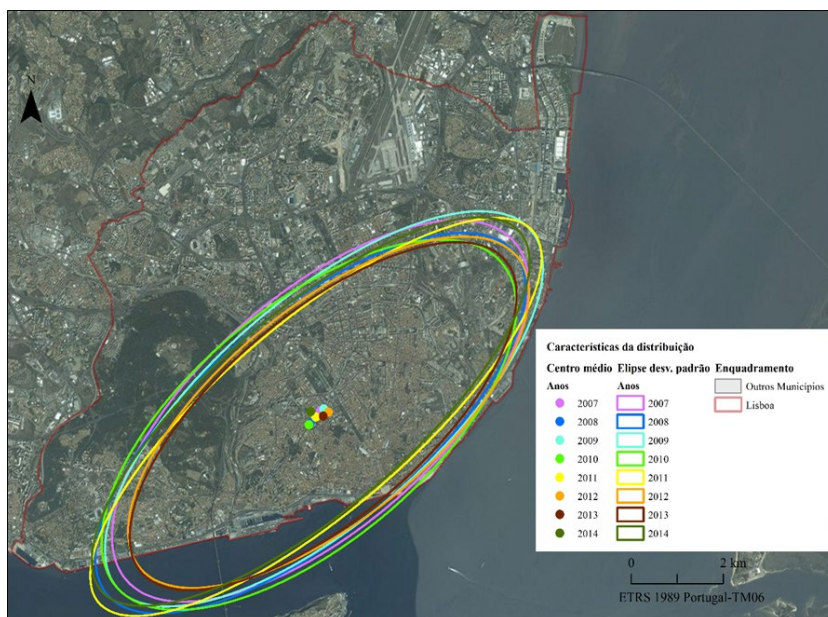
as 12 variáveis selecionadas através da análise do VIF (que apresentaram valores inferiores a 7,5), obteve-se o conjunto final de 7 variáveis.

## Resultados

### *Centro médio e elipse do desvio padrão*

A localização dos centros médios evidencia a tendência geral das observações de se afastarem do interior, situando-se mais para o quadrante sul - sudoeste da cidade.

As elipses (Figura 3) apresentam uma maior cobertura do litoral do que do interior, enfatizando a importância do eixo ribeirinho e das atrações situadas nas suas proximidades (e.g., Praça de Comércio, Mosteiro dos Jerónimos, Padrão dos Descobrimentos, etc.). As elipses expandem-se mais para o sudoeste do que para o nordeste, chegando a envolver, nessa direção, a área circunvizinha ao Padrão dos Descobrimentos. De referir que nenhuma das elipses consegue abarcar, no lado oposto, os pontos de interesse do Parque das Nações, o que demonstra a disparidade existente quanto ao número de observações localizadas em ambos os extremos.



**Figura 3.** Características da distribuição, Centro médio e Elipse do desvio padrão.

### Índice do vizinho mais próximo

O resultado do índice confirma que a distribuição espacial das observações é concentrada (sendo um padrão estatisticamente significativo,  $p\text{-value} = 0,000$  e com  $z\text{-score}$  inferior a  $-2,56$ ) e que, como tal, não é produto de um processo espacial aleatório.

### K-function (Multi-distance spatial cluster analysis)

Nos gráficos da função  $L(d)$ , para os anos 2007 e 2012 (Figura 4), é visível que a distribuição das observações é constantemente concentrada ao longo dos intervalos de distância estabelecidos (dos 50 m até os 1.300 m). Em todas as séries de distâncias, o padrão de *clustering* é estatisticamente significativo. As linhas dos valores  $L(d)$ , em ambos os casos, tendem a manter-se em constante crescimento até aos 1.150 m aproximadamente, indicando que após esta distância a significância estatística do padrão já é menos forte.

Os resultados confirmam que a distribuição espacial das observações é concentrada e que o padrão é estatisticamente significativo. Como se pode observar nos gráficos, a tendência de *clustering* é significativa, mesmo nas distâncias mais curtas, o que dá uma ideia sobre a estrutura da distribuição das fotografias. Em síntese, pode se concluir que as fotografias dos visitantes tendem a se concentrar em áreas de pequena dimensão e que essas áreas estão próximas umas das outras.

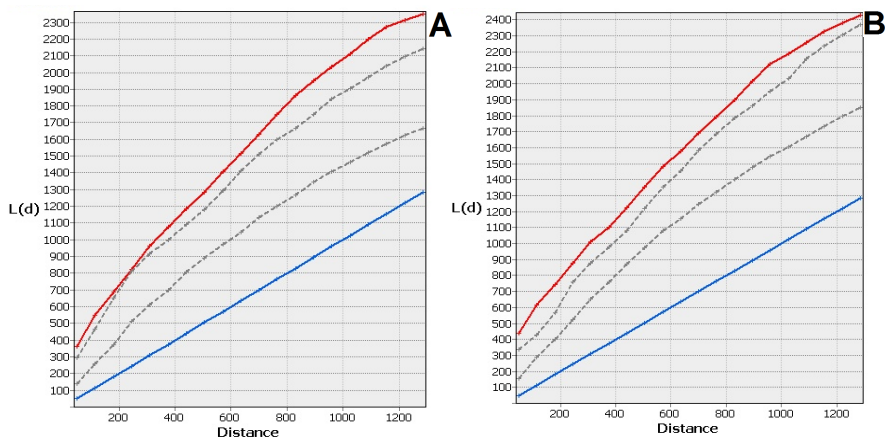


Figura 4. Função  $L(d)$  para as fotos agregadas dos anos: a) 2007; b) 2012.

### **Análise de Clusters e Outliers**

Como era de esperar, os *clusters* estão localizados nas zonas turísticas da cidade, e na circunscrição das atrações mais relevantes. Olhando para o mapa (Figura 5), o facto mais saliente tem a ver com que a quase totalidade dos *clusters* espaciais são de categoria *High-High*, ou seja, são locais onde valores elevados estão rodeados por valores semelhantes (elevados). Portanto, tem-se uma grande afluência de visitantes, tanto nesses locais como nas suas áreas contíguas.



**Figura 5.** Clusters e outliers espaciais do turismo em Lisboa.

Com efeito, verifica-se que as atrações turísticas funcionam como foco dos *clusters* espaciais. É sobre elas que se formam os *clusters*, e é a partir delas que estes se expandem para as áreas próximas. Em muitos casos, a extensão dos *clusters* segue a forma geométrica das atrações (*e.g.*, praças, passeios pedestres), ou dos passeios pedestres, não se expandindo além do perímetro deles.

Os *clusters* estão localizados em três grandes zonas, no Sul, na freguesia de Belém, no Centro, nas freguesias de Santa Maria Maior e Santo António e, no Norte, no Parque das Nações. Na parte sul, existem três *clusters* sobre monumentos: o da Torre de Belém, o do Padrão dos Descobrimentos e o do Mosteiro dos

Jerónimos. No centro, identificam-se 13 grupos de *clusters*, no Parque Eduardo VII, no Marquês de Pombal, nas praças dos Restauradores e Dom Pedro IV, no miradouro de São Pedro de Alcântara (incluindo a estação norte do Elevador da Gloria), no Elevador de Santa Justa e Convento do Carmo, na praça Luís de Camões (incluindo o Largo de Chiado), na estação norte do Elevador da Bica, no Convento da Graça (incluindo o miradouro da Graça), no Castelo de São Jorge, nos Miradouros das Portas do Sol e de Santa Catarina, na Igreja Sé de Lisboa e na praça do Comércio (incluindo parte da rua Augusta). E, no Norte, três *clusters*, dentro do sector do Parque das Nações.

### ***Análise de pontos quentes emergentes***

Quando a análise também é suportada pelo variável tempo, denotam-se ligeiras mudanças no esquema dos locais de concentração, que decorrem, obviamente, da junção das duas componentes (espacial e temporal). Ao analisar os resultados da APQE, em comparação com os da ACO, merece destaque o surgimento de vários locais de concentração e o desaparecimento de outros.

Olhando para o mapa (Figura 6), identificam-se cinco locais de concentração que não tinham sido descobertos no ponto anterior. Três deles estão fora das áreas com maior potencial turístico da cidade. De Sul para Norte, o primeiro local situa-se sobre o Museu da Eletricidade, e corresponde a um ponto cuja relevância aparece, nomeadamente, nos períodos mais recentes. Dentro do parque florestal de Monsanto, no antigo restaurante Panorâmico de Monsanto, desvela-se o segundo *hot spot*, cujo padrão temporal é de tipo consecutivo. Mais para o interior, sobressai o aglomerado sobre o Jardim Zoológico de Lisboa, cobrindo a quase totalidade do complexo e com uma relevância que permanece ao longo do período em análise. Na freguesia da Estrela, localiza-se um outro aglomerado, associado à Basílica da Estrela, com significância estatística apenas no último passo de tempo. Na freguesia de Santo António encontra-se o último destes novos focos, sobre o jardim do Torel, envolvendo a área do miradouro e constituindo um ponto quente nos intervalos de tempo finais (*hot spot* recente).

Por outro lado, os locais de concentração que deixaram de ser significativos, no contexto espaciotemporal, encontram-se no centro da cidade. São cinco, na freguesia de Santa Maria Maior (Miradouros das Portas do Sol e de Santa Catarina, igreja Sé de Lisboa, praças dos Restauradores e Dom Pedro IV, e Elevador de Santa Justa) e um, na freguesia da Misericórdia (Elevador da Bica). Nestes locais a significância parece não sobressair ao longo do tempo. Por exemplo, como se pode observar na visualização 3D (Figura 7), os *clusters* associados à Sé de Lisboa, em muitos intervalos de tempo, chegam apenas a um nível de confiança de 90%, sendo a sua significância estatística sem continuidade de relevo.

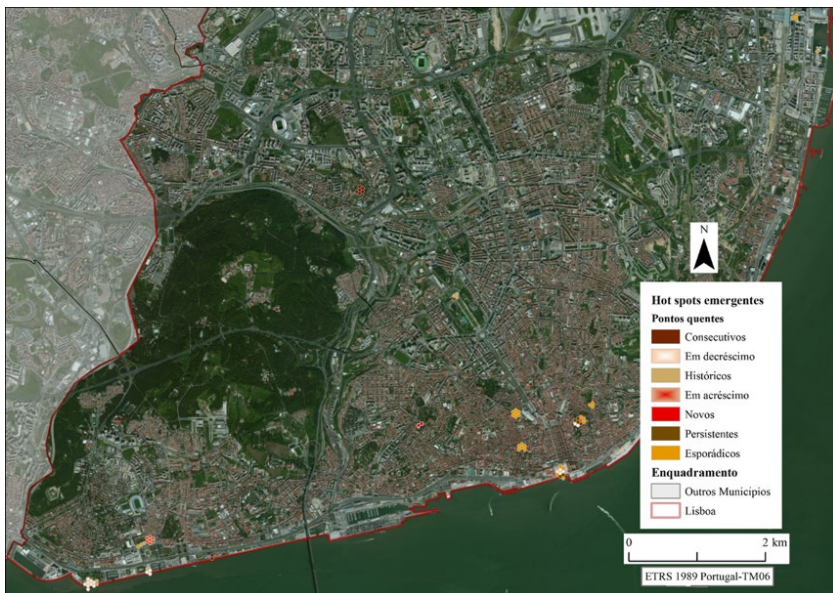


Figura 6. Tipos de pontos quentes emergentes.



Figura 7. Visualização 3D da conformação dos pontos quentes emergentes, zona centro.

De modo geral, as áreas turísticas típicas funcionam como um foco contínuo (ou consecutivo, segundo a nomenclatura utilizada) para os visitantes. No entanto, alguns lugares são mais prominentes, mantendo frequências cada vez mais elevadas de visitação turística nos últimos anos. Além disso, também é importante o aparecimento de outras áreas turísticas, além das fronteiras do núcleo tradicional de turismo da cidade, que emergem apesar de estarem espacialmente desconectadas mas que sobressaem pela sua relevância histórica.

### **Análise de regressão**

Da análise de regressão linear múltipla, retirou-se que os fatores mais importantes na explicação do padrão de concentração observado, são a proximidade a: Monumentos de Interesse Público (0,29); Docas de Recreio e Marinas (0,21); Arquitetura Premiada (0,15); Geomonumentos (0,12); Hotelaria (0,10); e Teatros (0,10). Como medidas de desempenho do modelo utilizaram-se os valores de  $R^2$  e  $R^2$  ajustado. Um valor de  $R^2$  ajustado de 0,68 indica que o modelo explica cerca de 68% da variação do fenómeno (*i.e.*, distribuição espacial dos visitantes dentro da cidade).

No entanto, para além da análise dos fatores e da predição do modelo, o que é também interessa em termos de planeamento turístico corresponde à análise dos resíduos do modelo. Na prática, analisando a Figura 8, o que os resíduos indicam é que, face às condições observadas, *i.e.*, fatores explicativos, há zonas em que o número de fotos é superior ao expectável (resíduos positivos) e outras onde há menos fotos do que seria previsível. São precisamente estas áreas, que podem necessitar de intervenções no âmbito do planeamento para que possam explanar todo o seu potencial.

No mapa dos resíduos, para além de Belém sobressair com valores acima do que o seu potencial justificaria, destaca-se de forma inesperada a zona da Baixa, com resíduos negativos, encontrando-se num nível inferior do que potencialmente poderá atingir. A zona do Parque das Nações segue a mesma tendência, mas não com um diferencial tão grande. Contudo, são claramente duas zonas identificadas que podem ser alvo de ações, tendo em vista a captação de mais visitantes.

### **Conclusões**

O uso de soluções tecnológicas nos processos de tomada de decisão contribuiu positivamente para avaliar as ações necessárias, o que é crucial para uma gestão sustentável do turismo e dos seus impactos. Com estas recentes soluções, potenciadas pelo envolvimento de ferramentas inteligentes, as cidades podem explorar novas abordagens e desafios em termos de planeamento do turismo, que se tornam acessíveis pelos dados disponibilizados em plataformas digitais online.



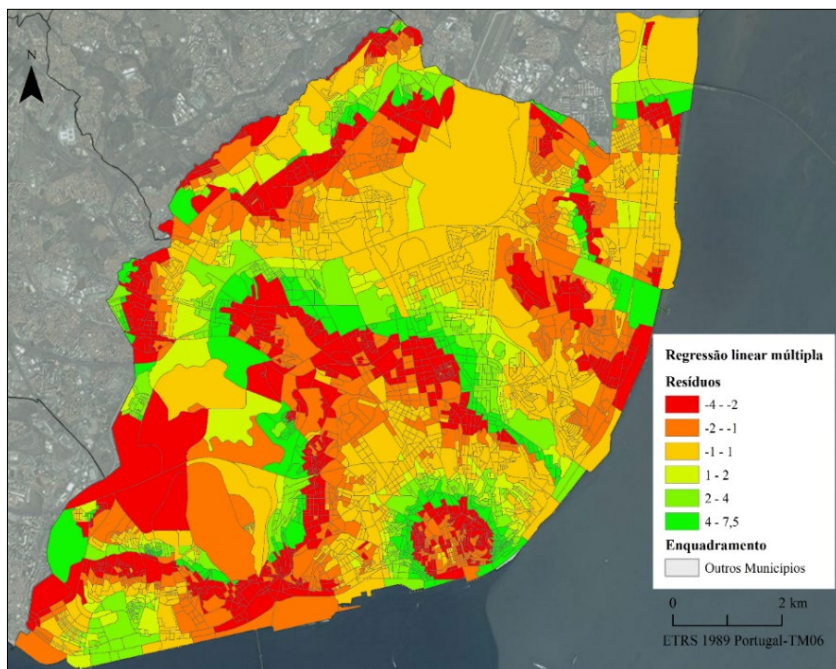


Figura 8. Resíduos do modelo de Regressão Linear Múltipla.

No caso de Lisboa, o mapeamento e conseqüente análise da pegada digital dos turistas levaram a constatar, por um lado, que a sua distribuição espacial não é homogênea e que, efetivamente, os turistas tendem a se concentrar nas áreas com maior potencial turístico da cidade, e, por outro lado, que são os elementos monumentais os que se mostram como focos de maior atratividade dos turistas, (*e.g.*, castelo, praças, igrejas).

O conjunto de dados (de 17.604 observações utilizadas para a análise) constitui uma amostra robusta que replica o espaço vivido pelos turistas. Os resultados da avaliação estatística da tendência geral de *clustering* evidenciou a sua aptidão para a análise espacial, deixando de lado qualquer possibilidade que sua gênese seja produto da aleatoriedade.

As áreas de interesse podem ser definidas analisando a concentração das fotografias dos turistas através de operações realizadas em SIG. Algumas delas podem nem ser consideradas atrações turísticas pelas autoridades competentes. Desta forma, podem saber quais as áreas que são mais atrativas para os visitantes e quais são as características dessas áreas (espaços verdes, monumentos, lugares com arquitetura moderna, etc.).

Esta evidência empírica sustenta o emprego destes dados, como um proxy sensível da distribuição espacial e temporal dos visitantes, para alcançar uma leitura quantitativa e geográfica sobre o comportamento da procura. As implicações dos resultados apresentados podem ser significativas para o planeamento e gestão das infraestruturas de transporte, o desenvolvimento de produtos e da imagem, e a gestão de impactos, especialmente para os destinos que são altamente dependentes do turismo.

## Bibliografia

- Ashworth, G. and Page, S. (2011). “Urban tourism research: Recent progress and current paradoxes”, em *Tourism Management*, 32(1), 1-15. DOI: 10.1016/j.tourman.2010.02.002.
- Anselin, L. (1995). “Local Indicators of Spatial Association —LISA”, *Geographical Analysis*, 27(2), 93-115, <<https://doi.org/10.1111/j.1538-4632.1995.tb00338.x>>.
- Buhalis, D. and Amaranggana, A. (2014). “Smart tourism destinations”, em Z. Xiang, & L. Tussyadiah (eds.), *Information and communication technologies in tourism 2014*, 553-564.
- Buhalis, D. and Law, R. (2008). “Progress in information technology and tourism management: 20 years on and 10 years after the internet - The state of eTourism research”, em *Tourism Management*, 29, 609-623. DOI: 10.1016/j.tourman.2008.01.005.
- Encalada, L.; Boavida-Portugal, I.; Ferreira, C. and Rocha, J. (2017). “Identifying tourist places of interest based on digital imprints: Towards a sustainable smart city”, em *Sustainability* 9(12), 2317. <<https://doi.org/10.3390/su9122317>>.
- ESRI (2016). *Emerging Hot Spot Analysis*. Retrieved May 28, 2018, from <<http://desktop.arcgis.com/en/arcmap/latest/tools/space-time-pattern-mining-toolbox/emerginghotspots.htm>>.
- Girardin, F.; Fiore, F.; Ratti, C. and Blat, J. (2008). “Leveraging explicitly disclosed location information to understand tourist dynamics: A Case Study”, em *Journal of Location Based Services*, 2(1), 41-56, <<http://dx.doi.org/10.1080/17489720802261138>>.
- García-Palomares, J.; Gutiérrez, J. and Mínguez, C. (2015). “Identification of tourist hot spots based on social networks: A comparative analysis of European metropolises using photo-sharing services and GIS”, em *Applied Geography*, 63, 408-417, <<https://doi.org/10.1016/j.apgeog.2015.08.002>>.
- Getis, A. and Ord, J.K. (1992). “The Analysis of Spatial Association by Use of Distance Statistics”, *Geographical Analysis*, 24(3), 189-206, <<https://doi.org/10.1111/j.1538-4632.1992.tb00261.x>>.

- Goodchild, M. and Li, L. (2012). “Assuring the quality of volunteered geographic information”, em *Spatial Statistics*, 1, 110-120. DOI: 10.1016/j.spasta.2012.03.002.
- Hayllar, B.; Griffin, T. and Edwards, D. (2008). “Urban Tourism Precincts: Engaging with the field”, em B. Hayllar, T. Griffin, and D. Edwards (eds.), *City Spaces-Tourist Places: Urban Tourism Precincts*, 3-18.
- Instituto Nacional de Estatística —INE (2015). “Estatísticas do Turismo 2014”. *Instituto Nacional de Estatística*, Lisboa, Portugal.
- Jin, C.; Cheng, J. and Xu, J. (2017). “Using User-Generated Content to Explore the Temporal Heterogeneity in Tourist Mobility”, *Journal of Travel Research*, 57(6), 779-791, <<https://doi.org/10.1177/0047287517714906>>.
- Kádár, B. (2014). “Measuring tourist activities in cities using geotagged photography”, em *Tourism Geographies*, 16(1), 88-104. DOI: 10.1080/14616688.2013.868029.
- Lau, G. and McKercher, B. (2006). “Understanding tourist movement patterns in a destination: A GIS approach”, em *Tourism and Hospitality Research*, 7(1), 39-49.
- Lo, I.; Mckercher, B.; Lo, A.; Cheung, C. and Law, R. (2011). “Tourism and online photography”, em *Tourism management*, 32, 725-731. DOI: 10.1016/j.tourman.2010.06.001.
- Munar, A. and Jacobsen, J. (2014). “Motivations for sharing tourism experiences through social media”, em *Tourism Management*, 43, 46-54. <<https://doi.org/https://doi.org/10.1016/j.tourman.2014.01.012>>.
- Pearce, D. (2001). “An integrative framework for urban tourism research”, em *Annals of Tourism Research*, 28(4), 926-946. PII S0160-7383(00)00082-7.
- Tammet, T.; Luberg, A. and Järv, P. (2013). “Sightsmap: Crowd-sourced popularity of the world places”, em L. Cantoni and Z. Xiang (eds.), *Information and Communication Technologies in Tourism 2013*, 314-325.
- Urry, J. and Larsen, J. (2011). “The tourists gaze 3.0”, London, SAGE Publications, Ltd.

