

Desenvolvimento de serviços e aplicações

Sistemas de recomendação, entender e melhorar o tráfego, melhorar o planejamento urbano...

Análise de dados urbanos

Analisar os resultados

Executar scripts

Editar scripts

Gerenciamento de dados urbanos

Coleta

Processamento

Armazenamento/
Modelagem

Fontes de dados urbanos



Desenvolvimento de serviços e aplicações

Sistemas de recomendação, entender e melhorar o tráfego, melhorar o planejamento urbano...

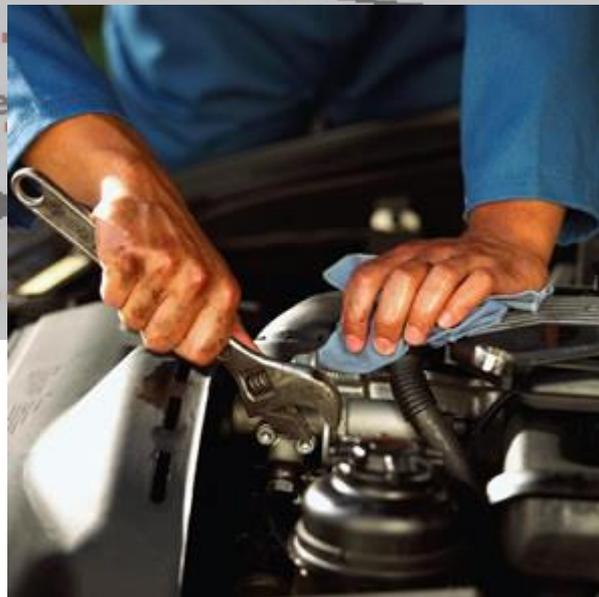
Análise de dados urbanos

Analisar os resultados

Executar scripts

Editar scripts

Fontes de dados urbanos



Gerenciamento

Armazenamento/Modelagem

Funcionamento
de cidades

Padrões sociais,
econômicos e culturais

Mobilidade
urbana

Detecção de Eventos e
Interesses

Problemas das
cidades

Funcionamento de cidades



City Image

Revealing the City that We Cannot See
ACM Transactions on Internet Technology - 2014



Belo Horizonte, Brazil

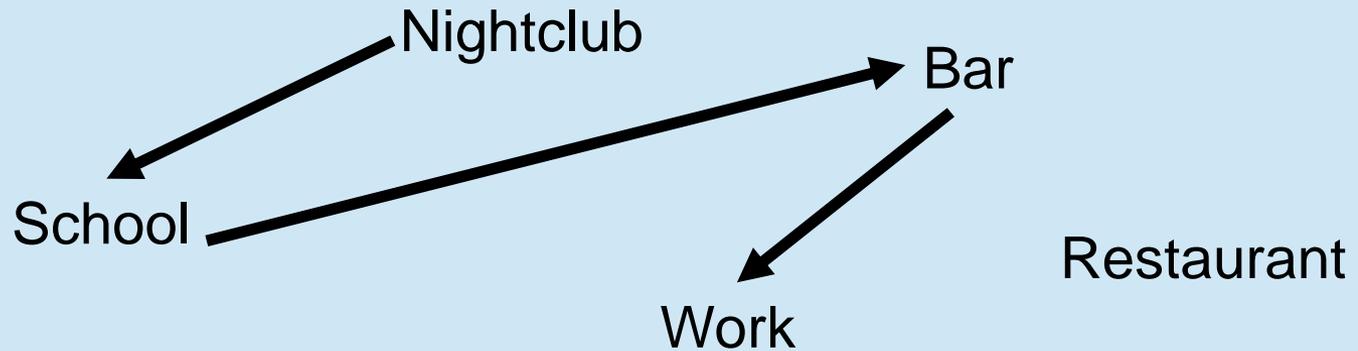


New York, USA



As transições são aleatórias?

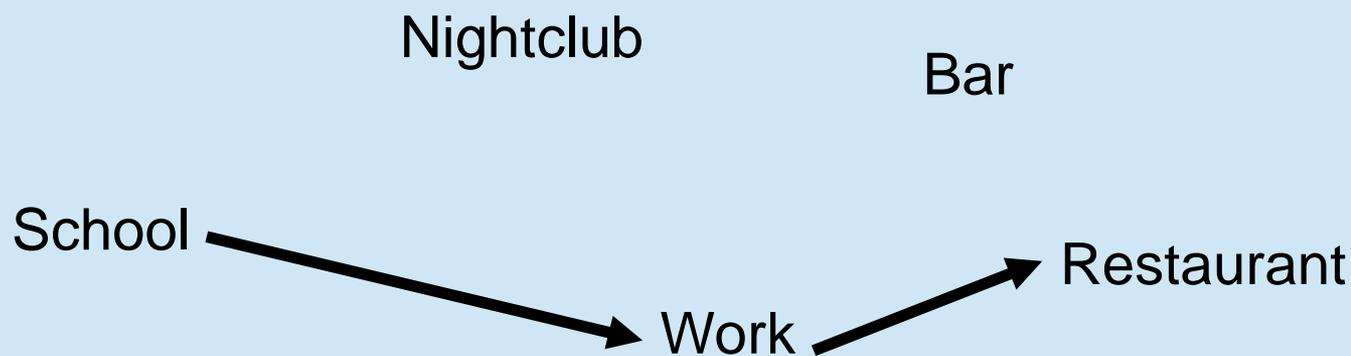
Segunda-feira típica





Existem transições que são mais favoráveis

Segunda-feira típica



Grafo de transição urbana:

User1

Tempo 1



Pizzaria

“Food”

Grafo de transição urbana:

User1

Tempo 1



Pizzaria
“Food”

User1

Tempo 2



Jazz bar
“Nightlife”

Grafo de transição urbana:



Grafo de transição urbana $G(V, E)$
Nós são as categorias de locais

Grafo de transição urbana:



Grafo de transição urbana $G(V, E)$
Peso de uma aresta é o número total de
ocorrências das transições

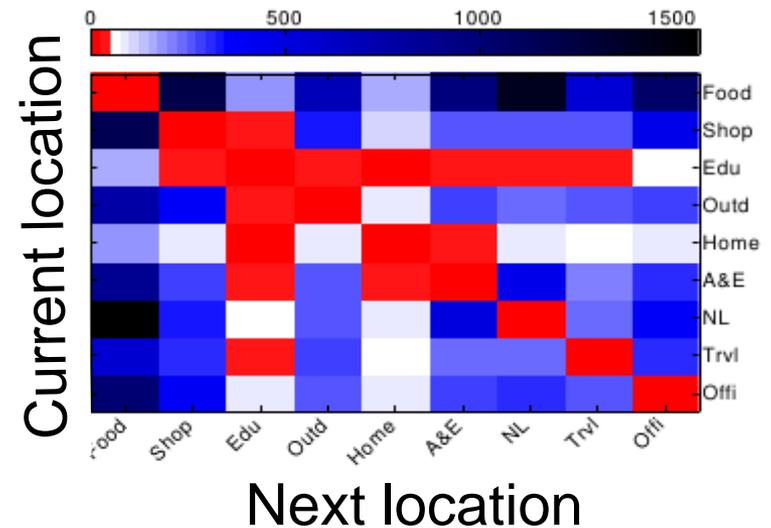
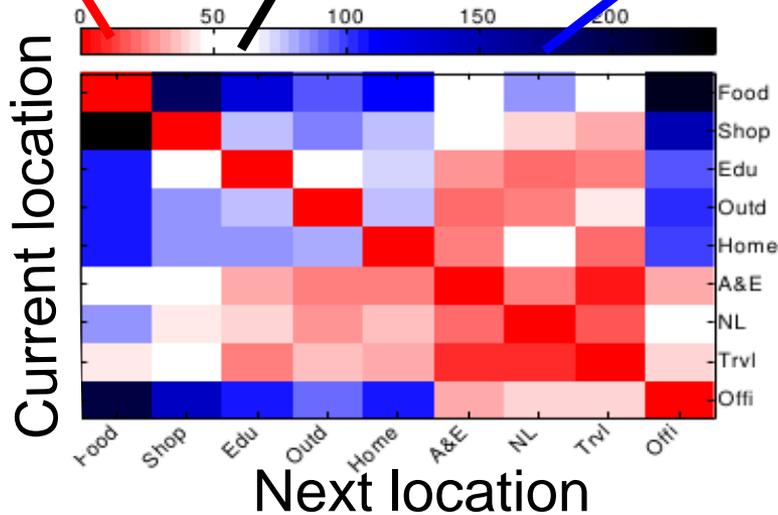
City Image



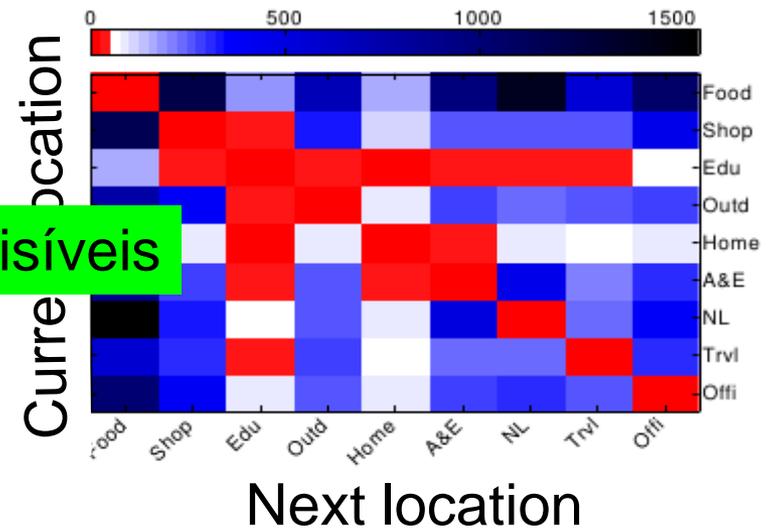
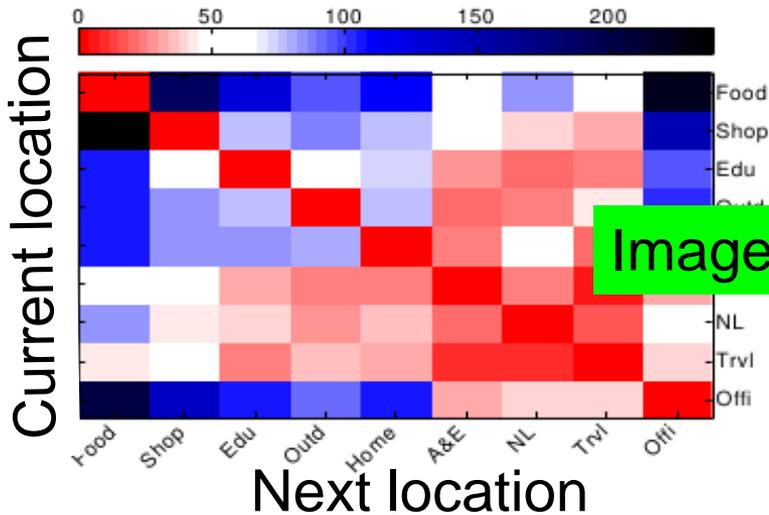
Rejeição

Indiferença

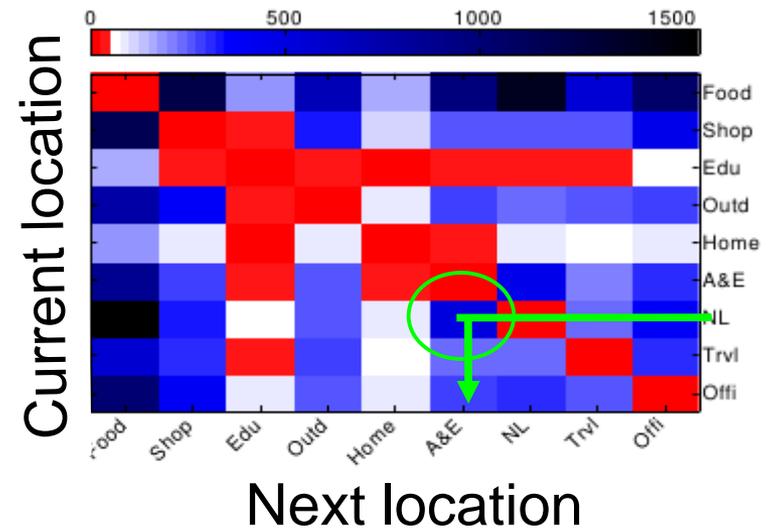
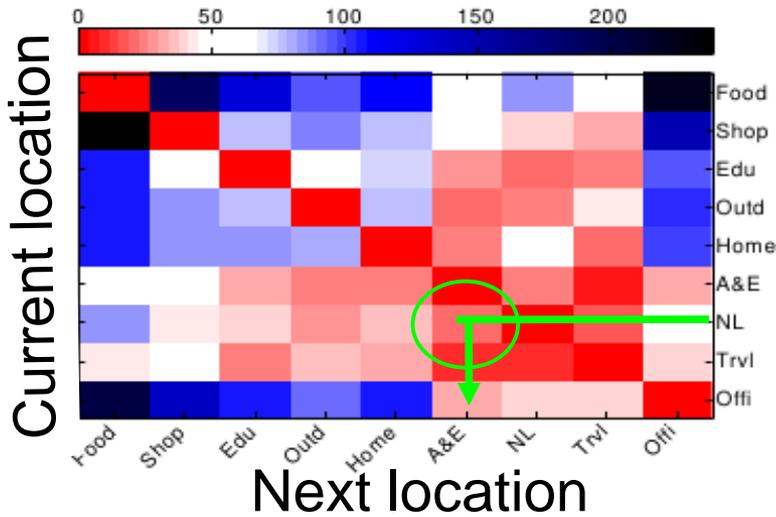
Favorável



City Image

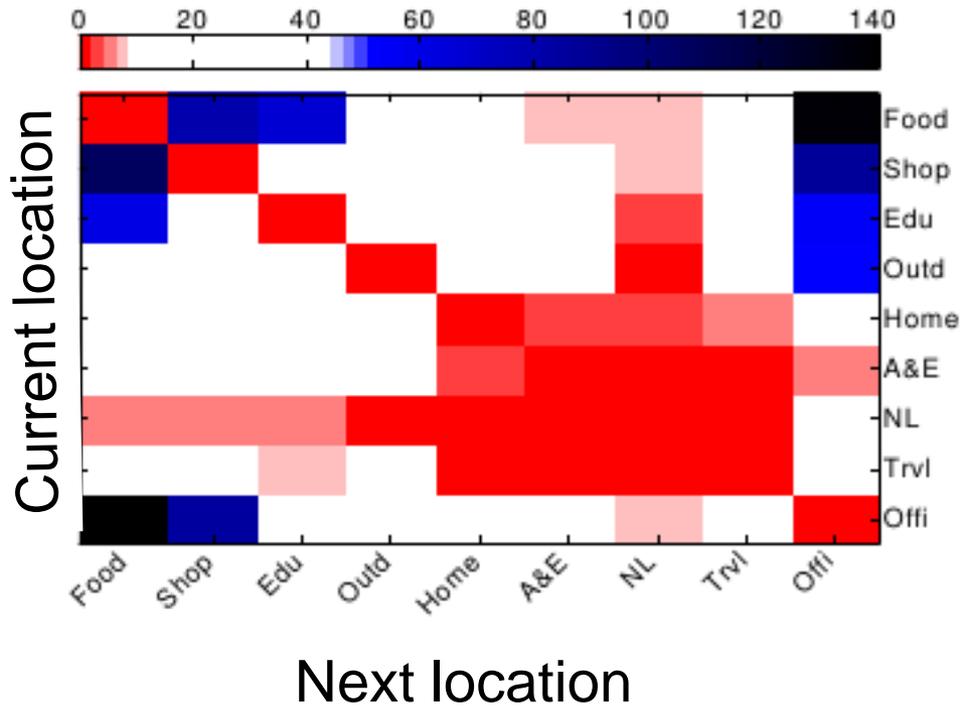


City Image

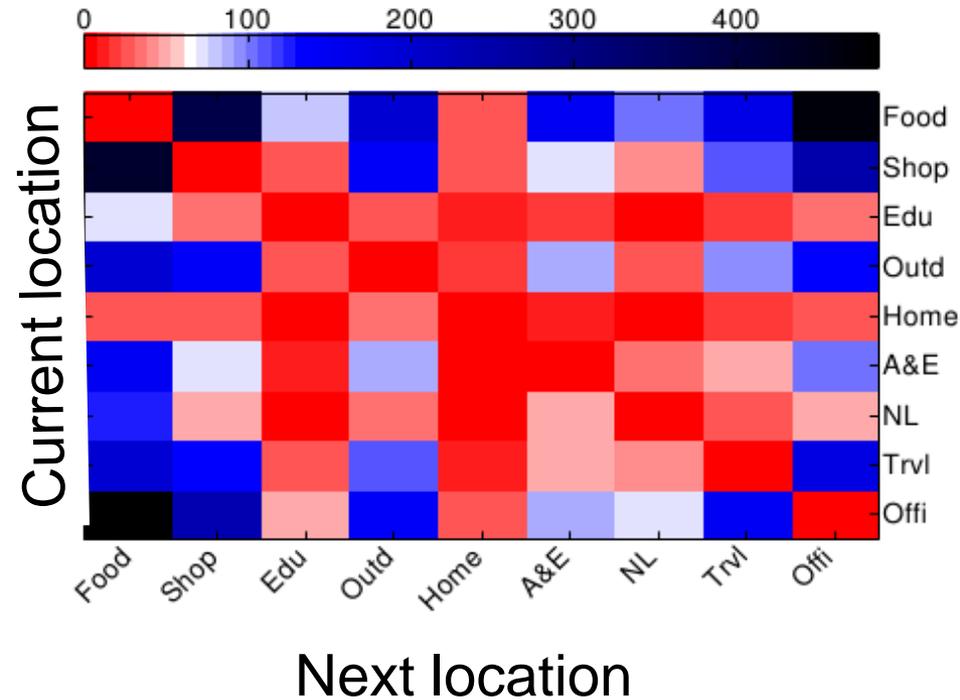


É possível diferenciar as rotinas dos habitantes

Belo Horizonte – weekday (DAY)

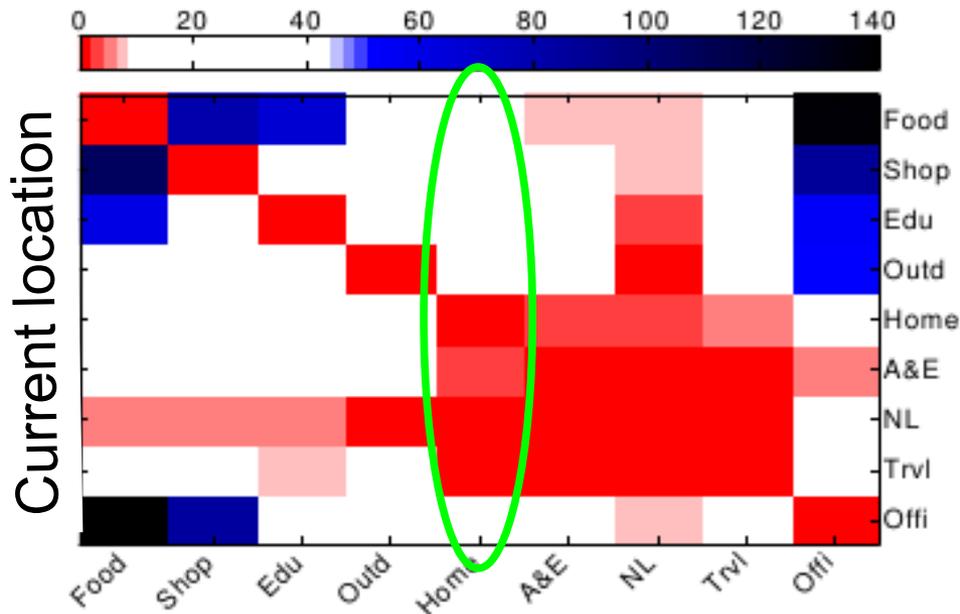


NY – weekday (DAY)



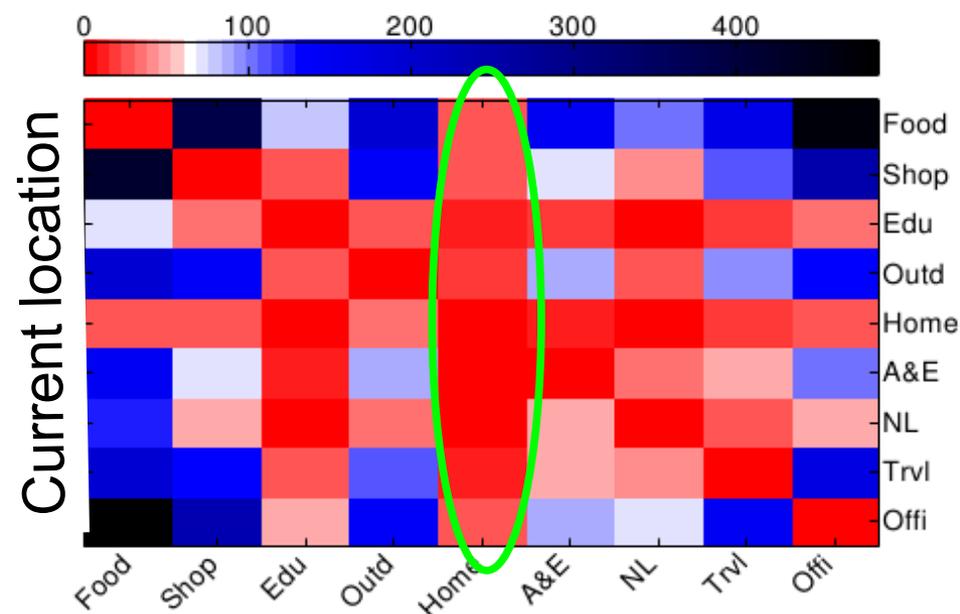
É possível diferenciar as rotinas dos habitantes

Belo Horizonte – weekday (DAY)



Next location

NY – weekday (DAY)

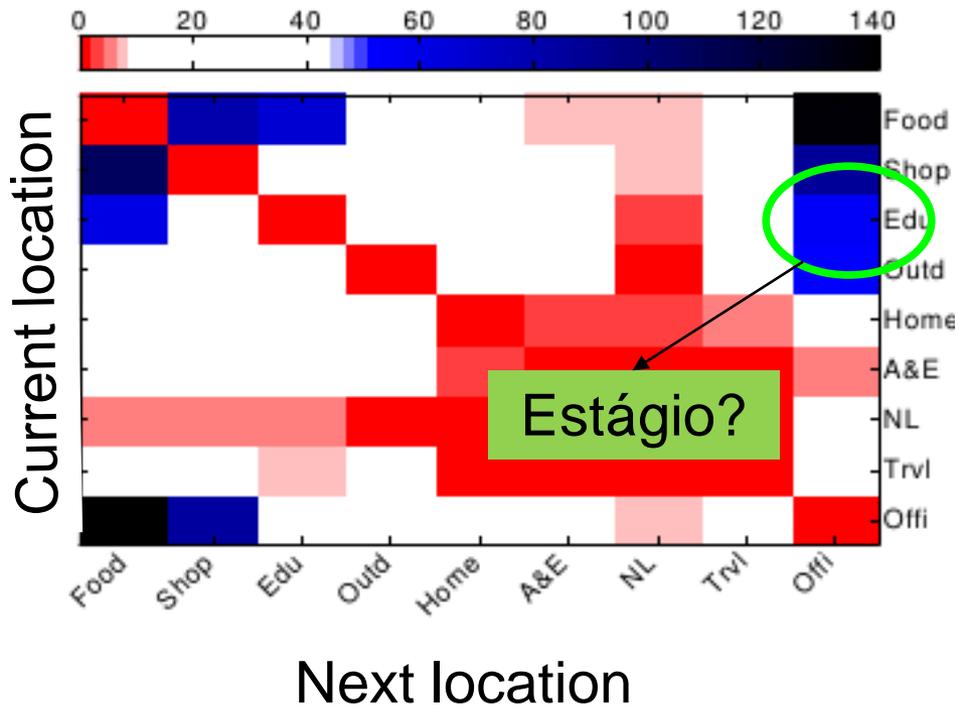


Next location

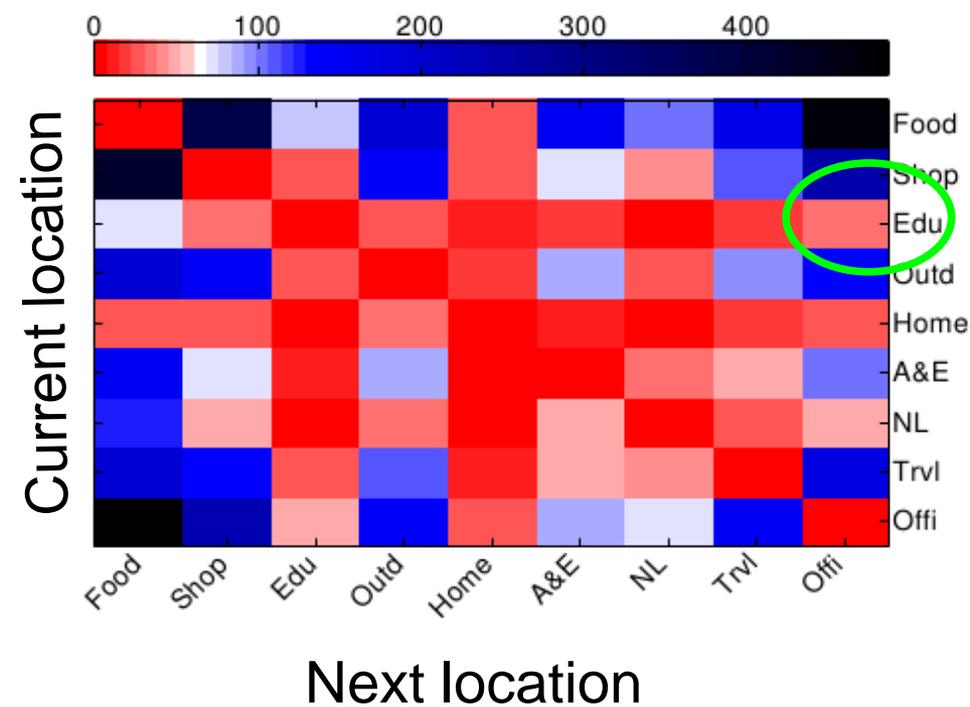
Não gostam de informar quando chegam em casa

É possível diferenciar as rotinas dos habitantes

Belo Horizonte – weekday (DAY)

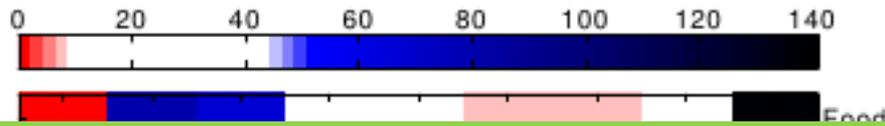


NY – weekday (DAY)

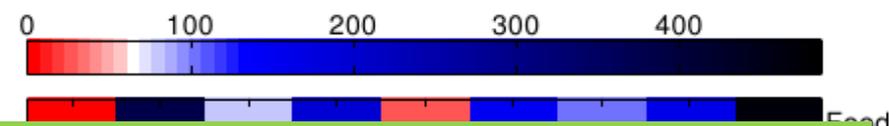


É possível diferenciar as rotinas dos habitantes

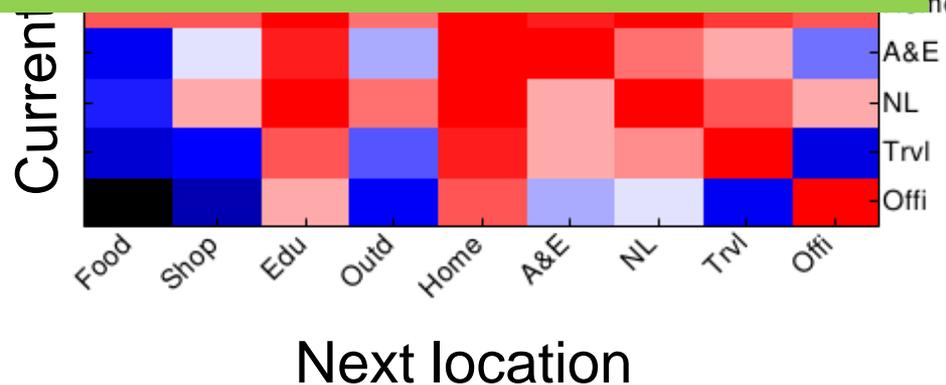
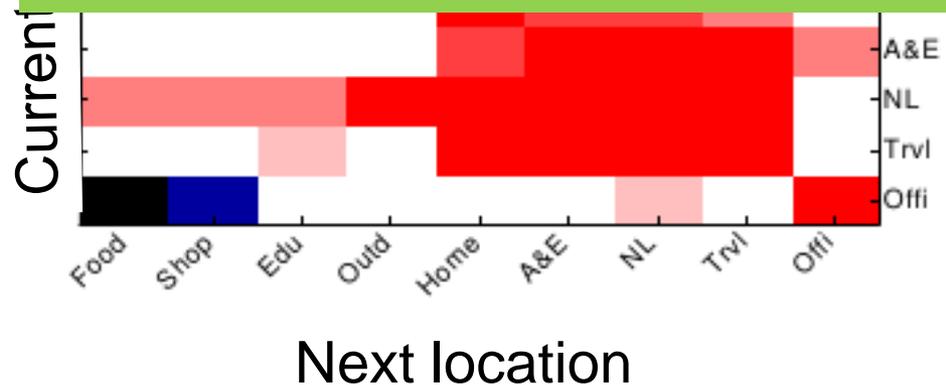
Belo Horizonte – weekday (DAY)



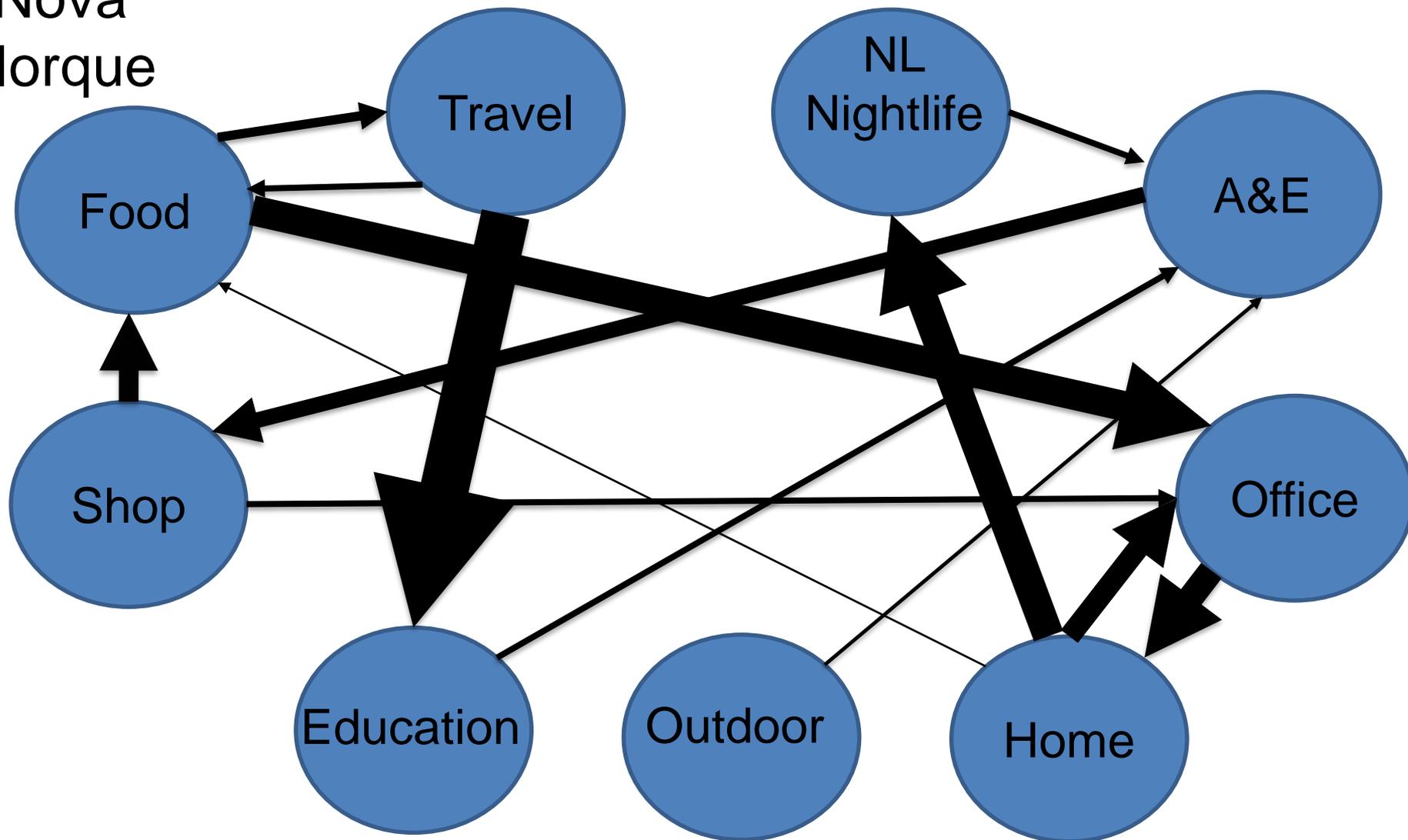
NY – weekday (DAY)



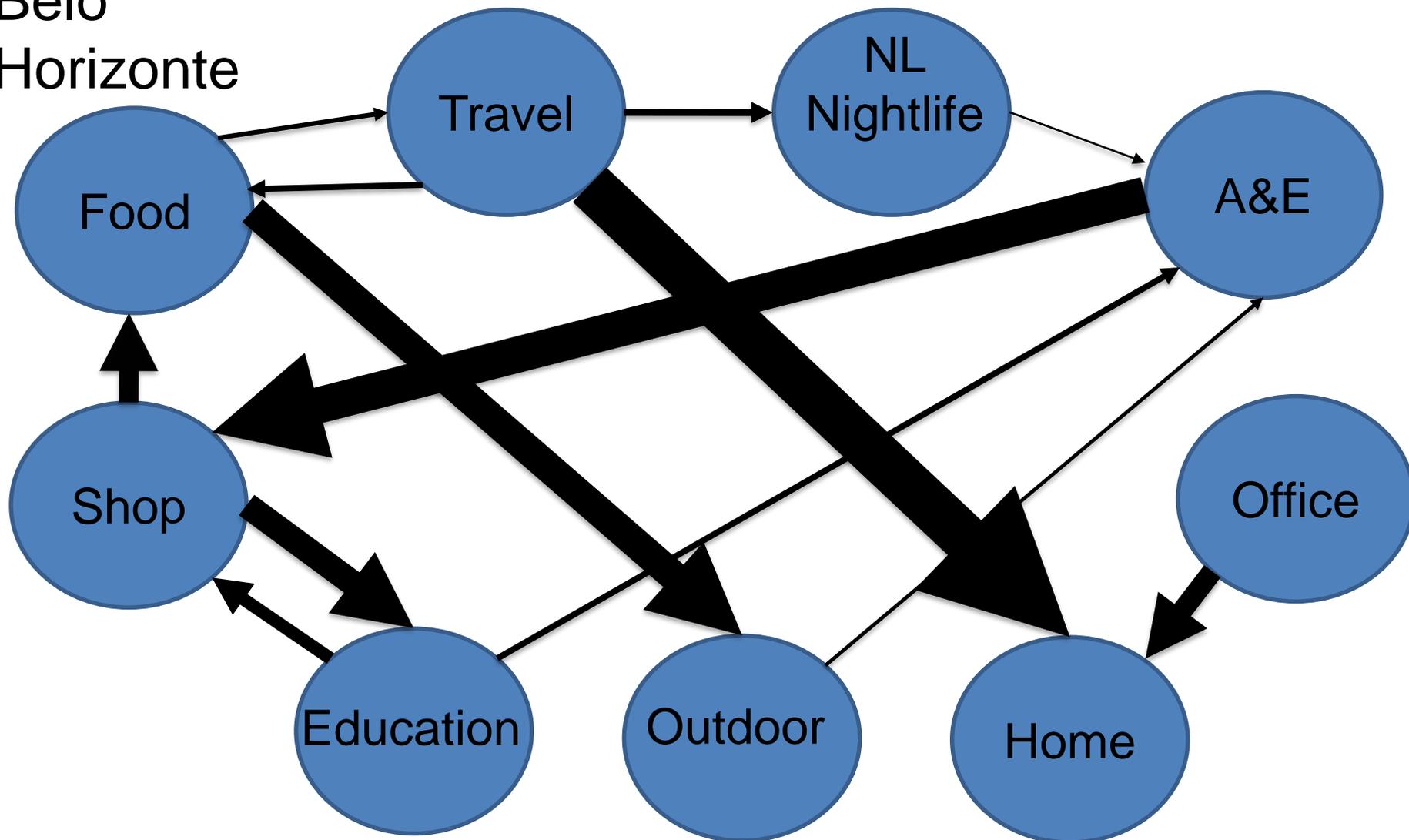
Agrupamento de cidades vão de acordo com o senso comum



Nova
Iorque



Belo Horizonte

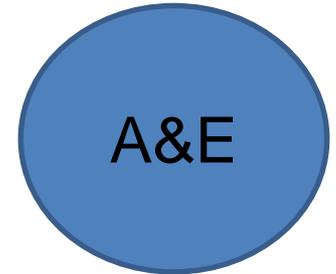
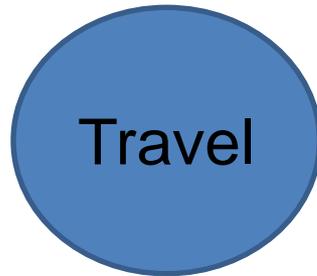
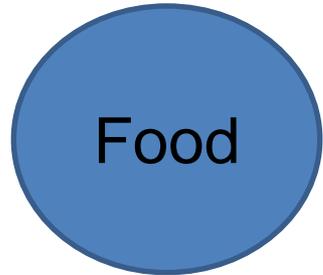


Vários grafos aleatórios

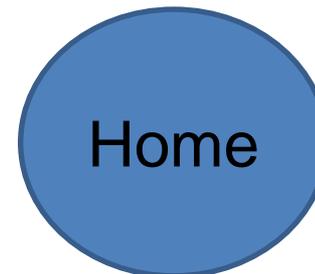
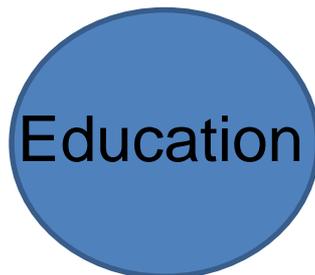
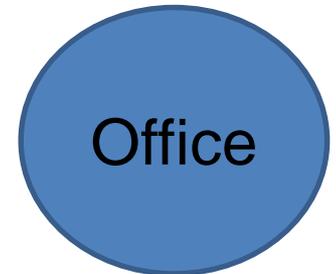
Mesmo número de nós do grafo original $G(V, E)$

Ao invés de considerar a transição atual escolhemos uma aleatória

Aleatório



Escolha aleatória de arestas



Distribuição dos pesos das arestas aleatórias -> normal $N(\bar{w}, \sigma_w)$

Usar a média \bar{w} e desvio padrão σ_w da distribuição

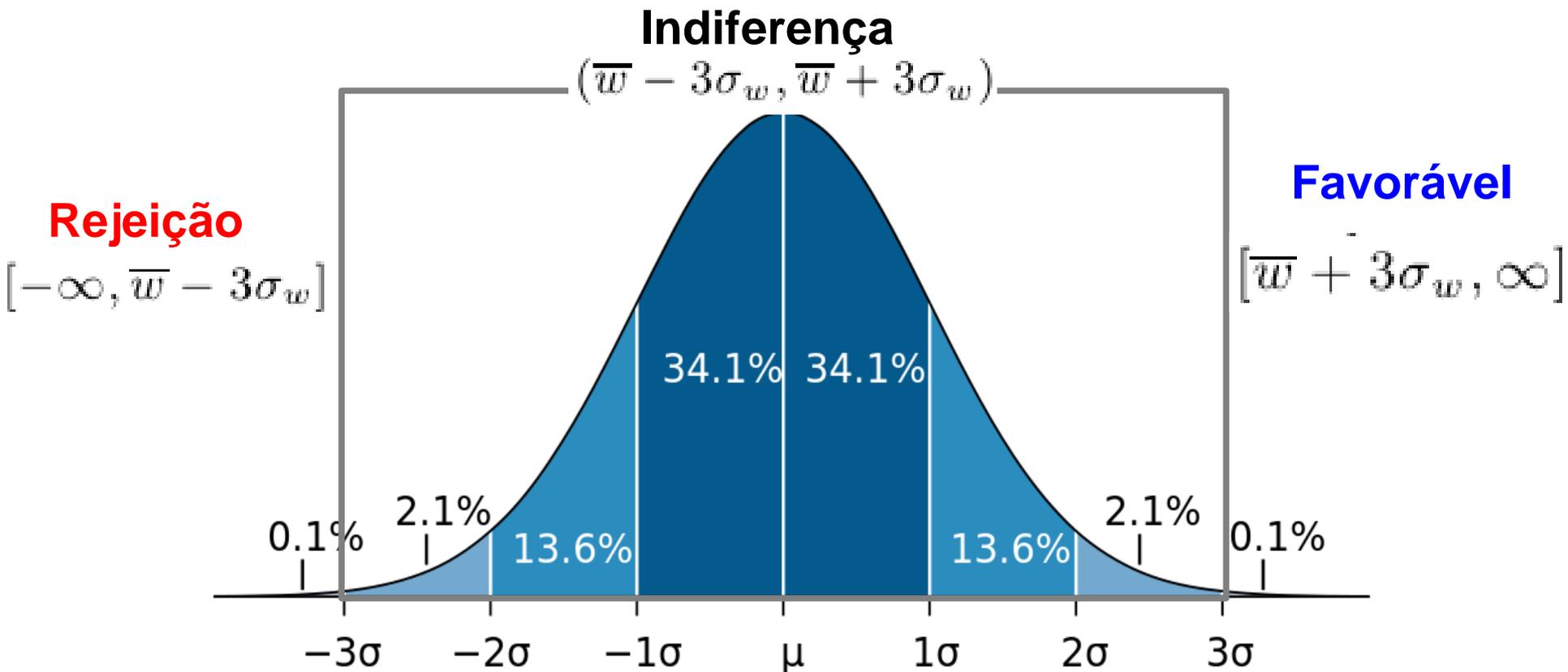


Table 4: Top 3 edge weights between sub-categories.

City	Edges
Belo Horizonte	University → Home (35), Office → Brazilian Restaurant (33), Neighborhood → Home (21)
Mexico City	Stadium → Concert Hall (86), Concert Hall → Stadium (56), Office → Mexican Restaurant (44)
New York City	Coffee Shop → Office (66), Office → Coffee Shop (44), Subway → Office (29)
São Paulo	Office → Brazilian Restaurant (120), Train Station → Subway (65), Subway → Office (58)
Paris	Subway → Office (12), Train Station → Office (12), Office → French Restaurant (12)
Tokyo	Train Station → Subway (573), Subway → Train Station (501), Train Station → Ramen or Noodle House (222)
Rio de Janeiro	Office → Brazilian Restaurant (57), Office → Restaurant (36), Gym → Home (28)
Sydney	Café → Office (16), Office → Café (9), Bridge → Café (5)

Livelihoods

Utilizing Social Media to Understand the Dynamics of a City
ICWSM'12

Hipóteses:

- Áreas não são definidas apenas por locais
- As pessoas que fazem parte da área também são importantes

- Identificar novos limites para os bairros que refletem padrões atuais de atividades coletivas

Pessoas similares
tendem a ir nas
mesmas áreas



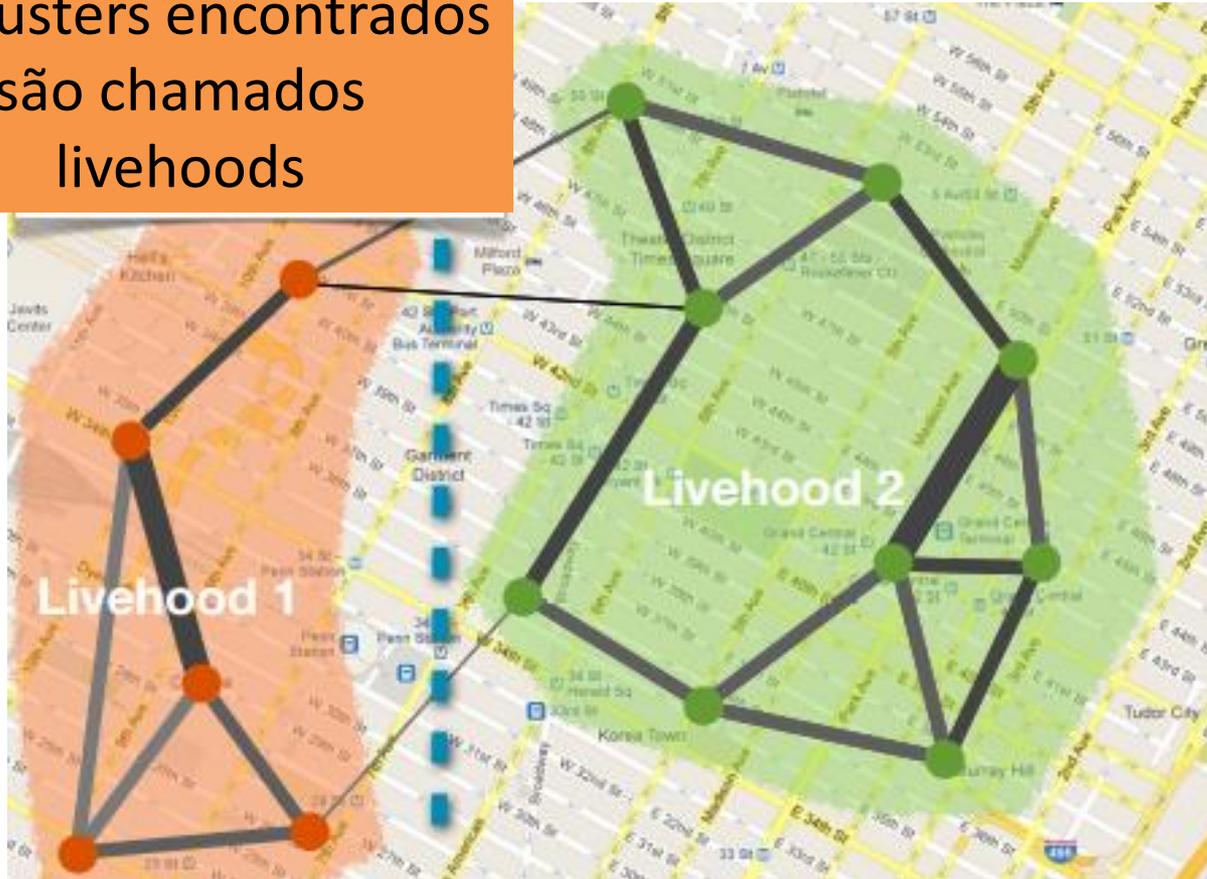
Usar este padrão para
computar a relação
entre locais



As relações podem ser usadas para identificar novas fronteiras

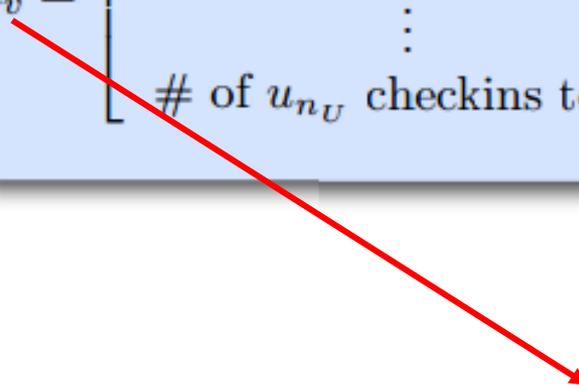


Os clusters encontrados
são chamados
livehoods



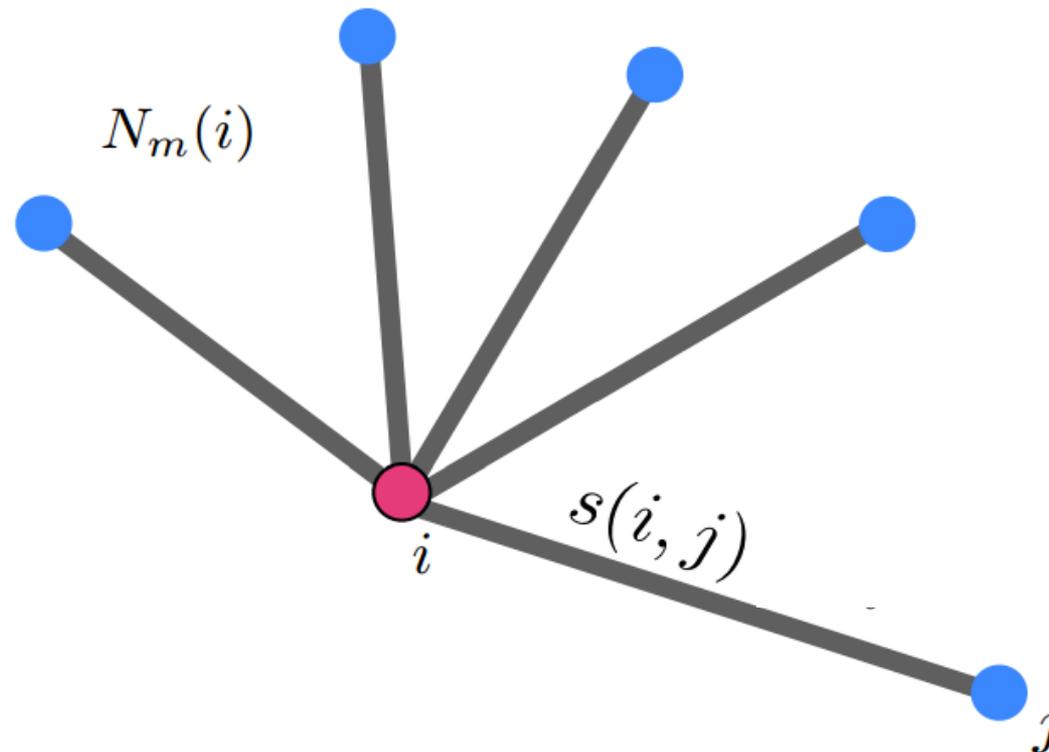
$$c_v = \begin{bmatrix} \# \text{ of } u_1 \text{ checkins to } v \\ \# \text{ of } u_2 \text{ checkins to } v \\ \vdots \\ \# \text{ of } u_{n_U} \text{ checkins to } v \end{bmatrix}$$

Afinidade social dos locais


$$s(i, j) = \frac{c_i \cdot c_j}{\|c_i\| \|c_j\|}$$

Cosine similarity

A ideia é encontrar clusters geograficamente próximos
(impor esse limite)



Conecta cada local (nó) aos
seus m vizinhos

O peso das arestas é a
afinidade social dos locais

Detecção de eventos e interesses



Pontos de interesse (PDI)

A picture of Instagram is Worth More than a Thousand
Words: Workload Characterization and Application.
DCOSS'13

Uma aplicação que emerge naturalmente é de identificação de pontos de interesse (PDI) em uma cidade

Cada foto representa, implicitamente, um interesse de um indivíduo em um determinado instante



Cada foto possui uma coordenada

(x,y)

Pontos de interesse



(x,y)



(x,y)

Calculamos a distância entre cada ponto



(x,y)

(x,y)

CLUSTER 1

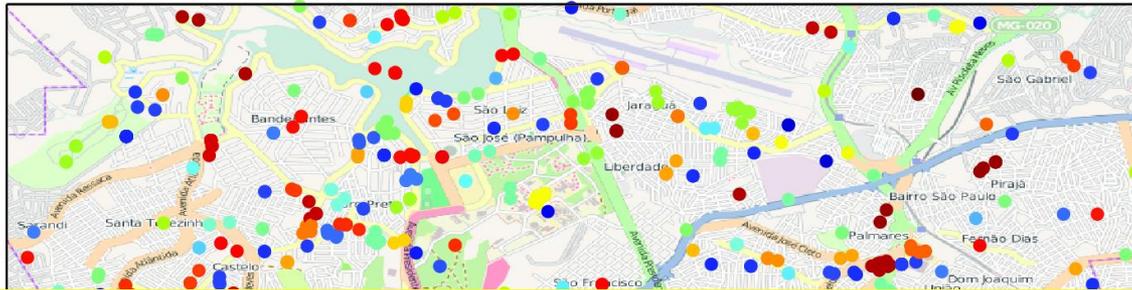
Agrupa-se fotos com distâncias inferiores a 250m

CLUSTER 2

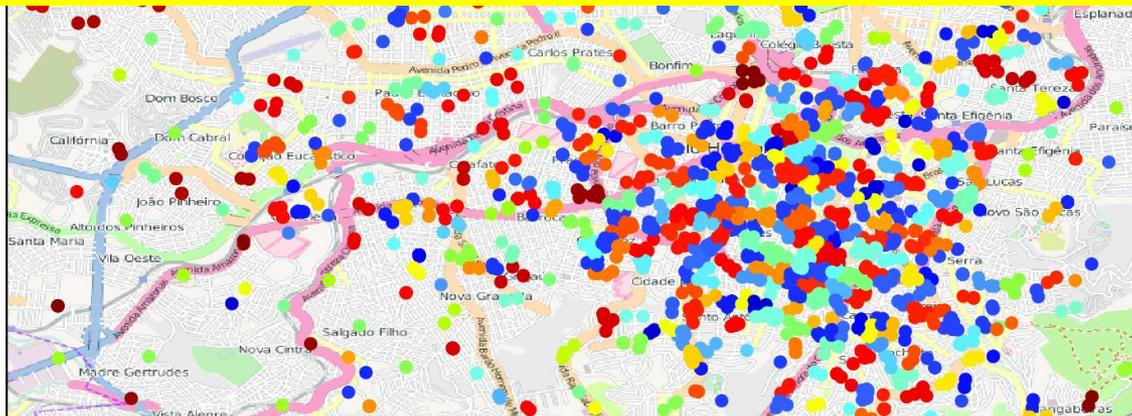


(x,y)

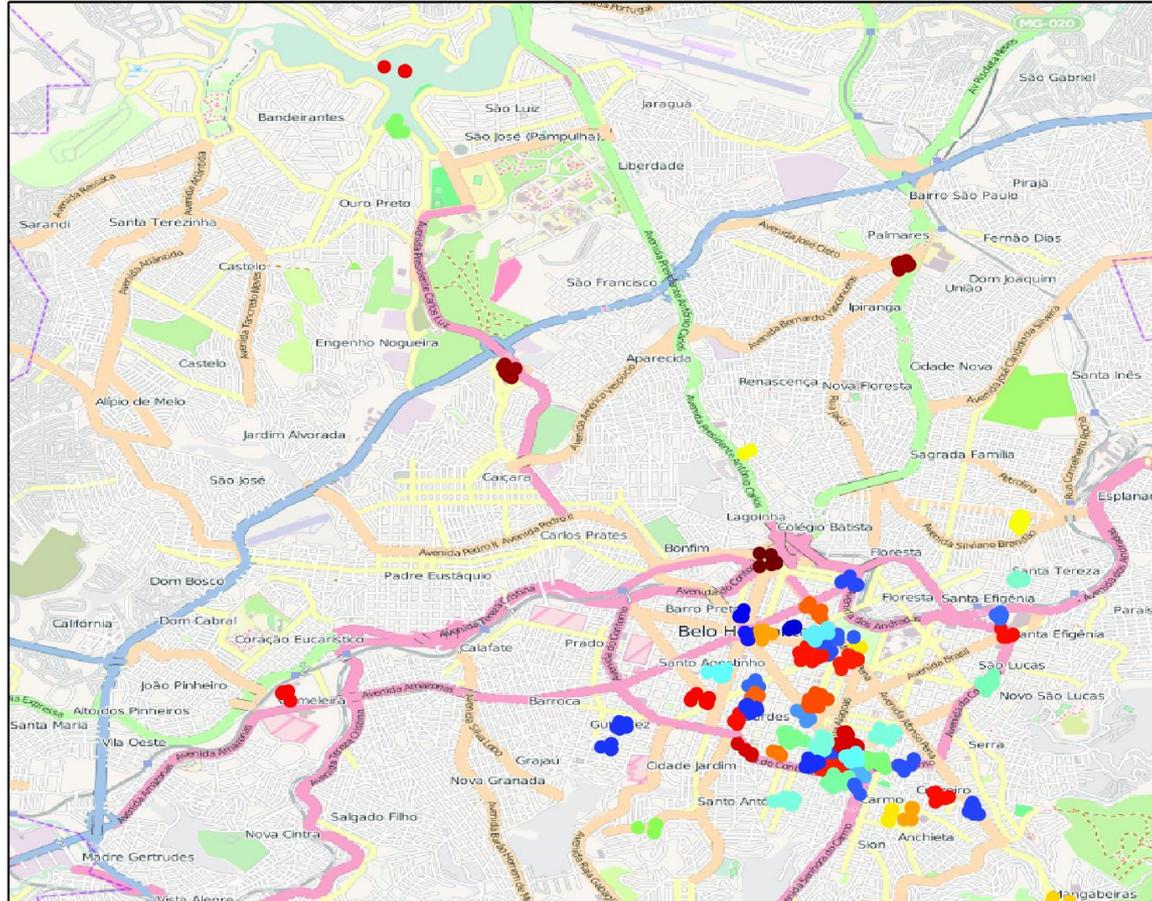
(x,y)



Identificação de áreas populares
Procedimento explicado em seguida



Todos os clusters



Pontos de interesse

Conjectura:

Grande parte dos turistas seguem uma trajetória bem conhecida dentro da cidade.

Em cada ponto turístico ele tira uma ou mais fotos e parte para o ponto turístico seguinte.

Grafos de transição urbana: $G(V, E)$

- **nó** $v_i \in V$ são todos os PDIs
- **aresta** (i, j) existe do nó v_i para o nó v_j se ocorreu uma transição de v_i para v_j
- **peso** $w(i, j)$ de uma aresta é o número total de transições que ocorreram do nó v_i para o nó v_j

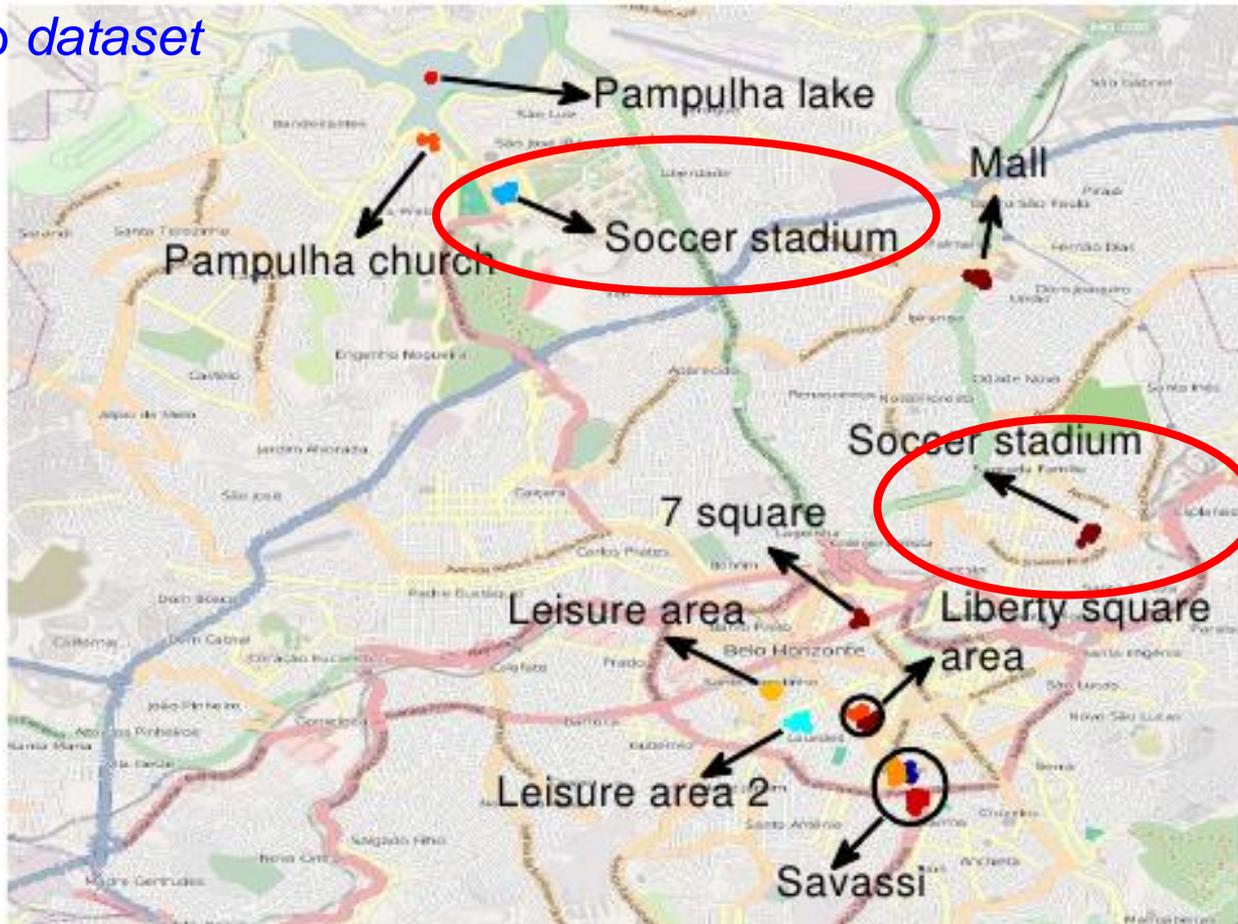
Separaremos transições que são populares



Pontos de interesse

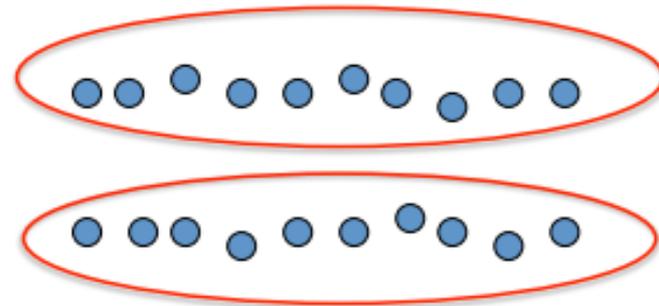
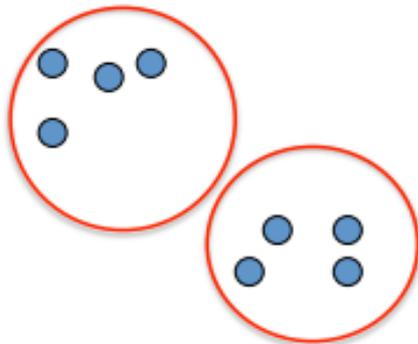
Futebol voltou!

Novo dataset



- **Agrupamento**

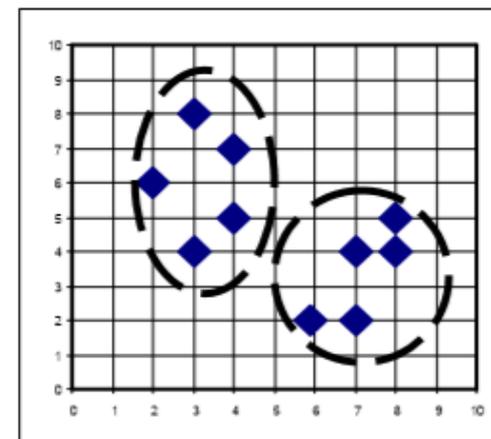
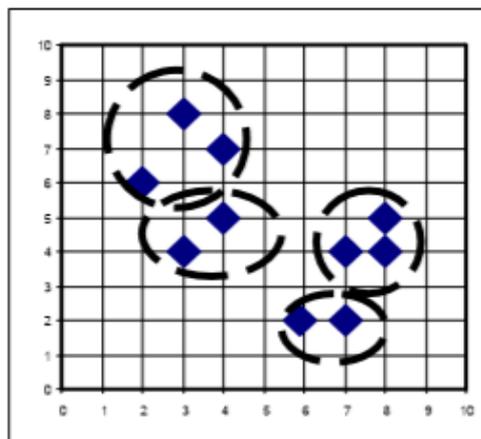
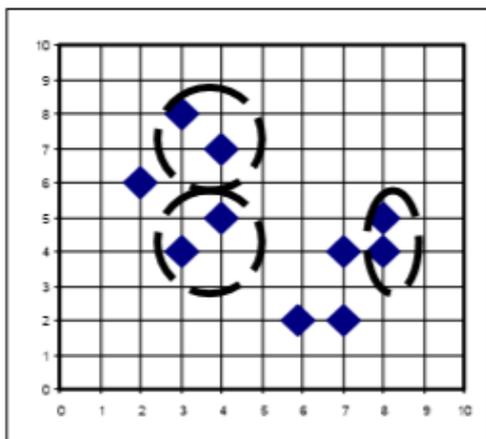
Ideia básica: agrupar instâncias similares



- **O que similar significa?**
 - **Contexto espacial:** distância geográfica (haversine)
 - Resultados do agrupamento dependem disso
- **Clusterização/agrupamento**
 - É uma classificação não supervisionada: sem classes predefinidas

Agrupamento hierárquico (aglomerativo)

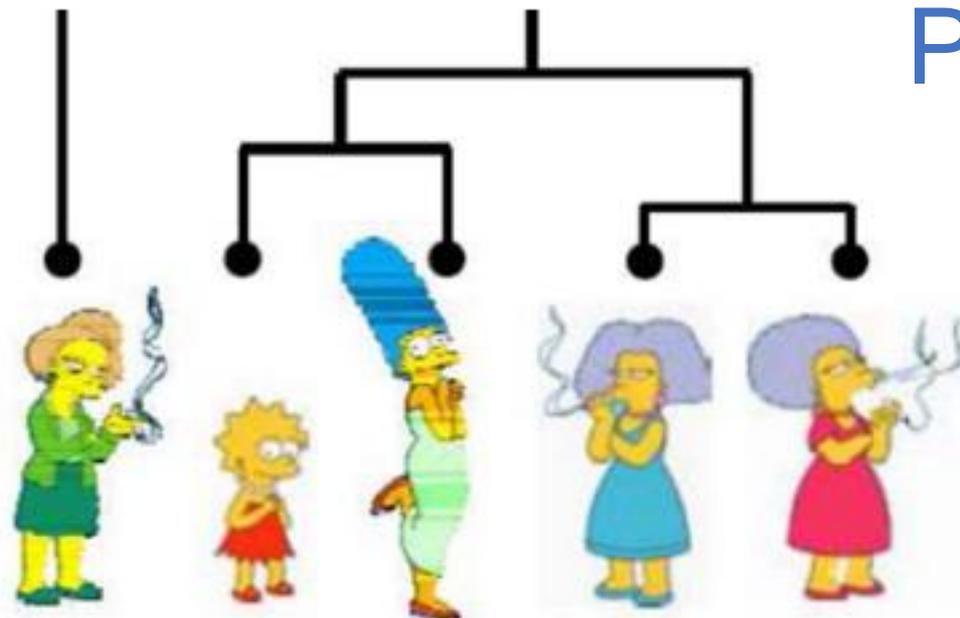
Aglomeración



Distância entre os pontos: geográfica (haversine)

Imagem de Stanley R M Oliveira

Agrupamento hierárquico (aglomerativo)

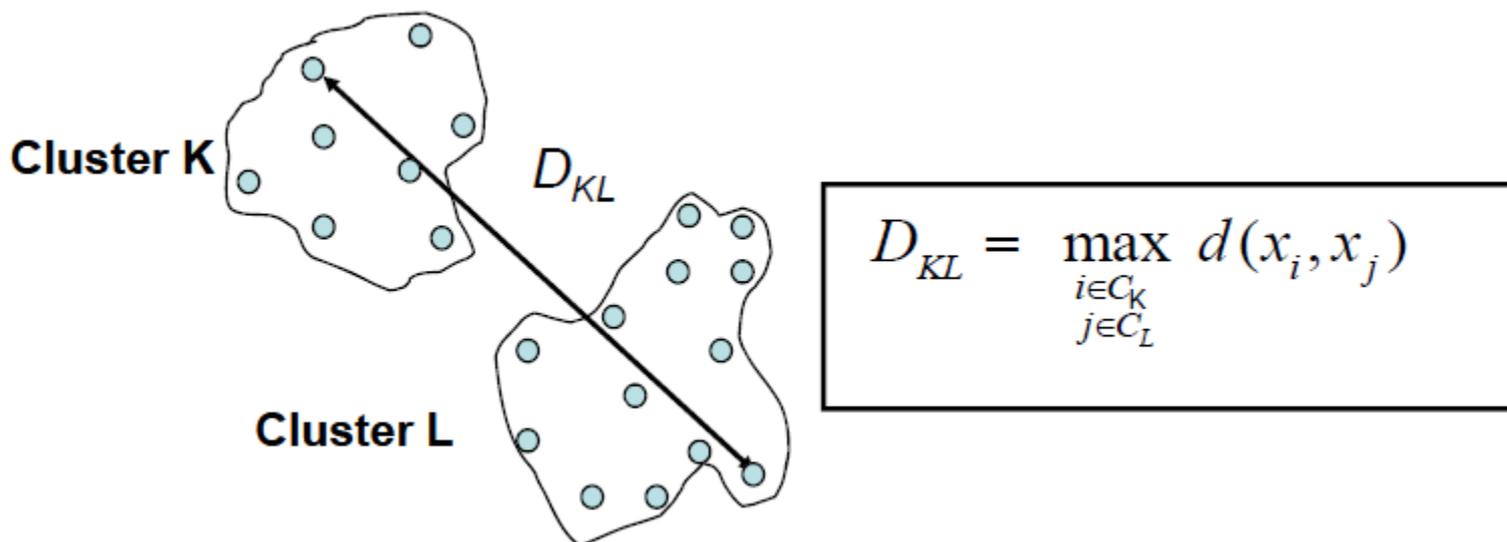


Parentesco

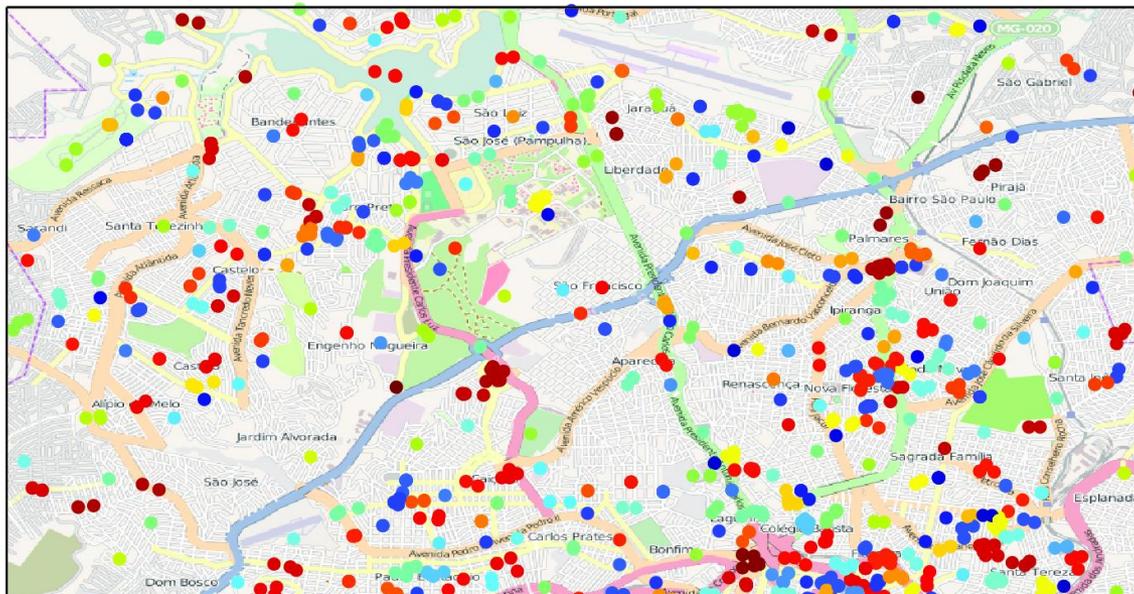
Imagem de David Sontag

Agrupamento hierárquico (aglomerativo)

Complete linkage – critério para aglomerar clusters



Identificação de PDIs

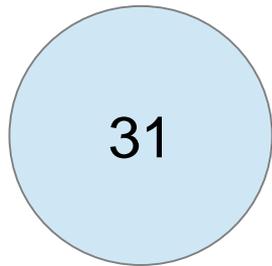


Para cada cluster é considerado apenas um dado por usuário

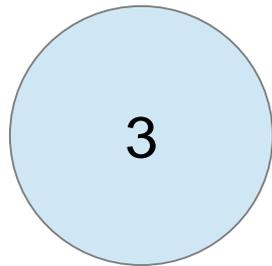


Todos os clusters

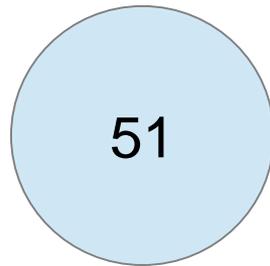
Clusters originais:



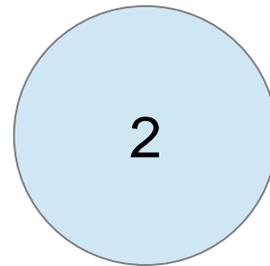
Cluster 1



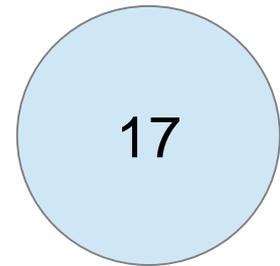
Cluster 2



Cluster 3

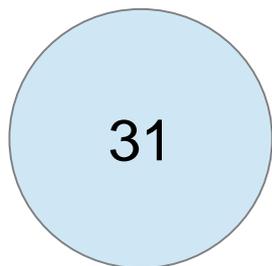


Cluster 4

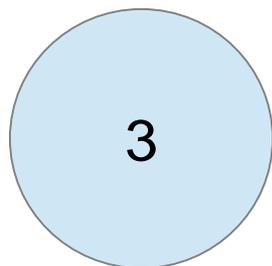


Cluster 5

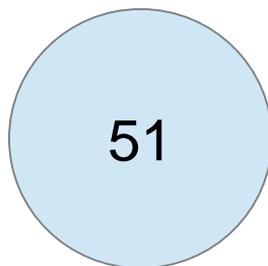
Clusters originais:



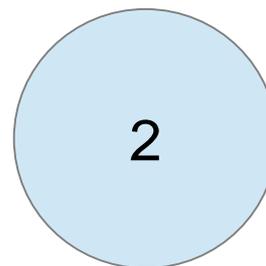
Cluster 1



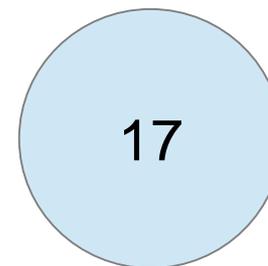
Cluster 2



Cluster 3

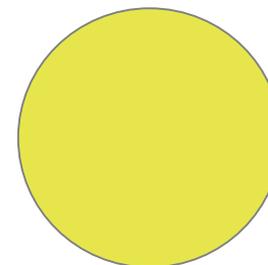
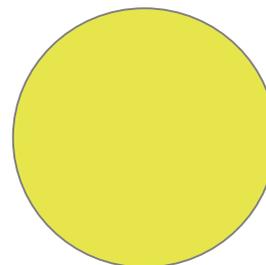
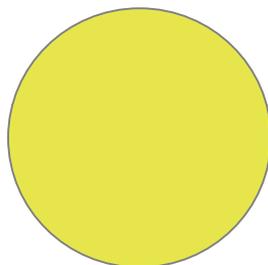
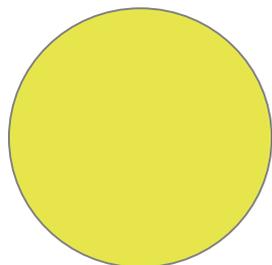
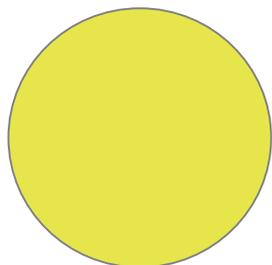


Cluster 4

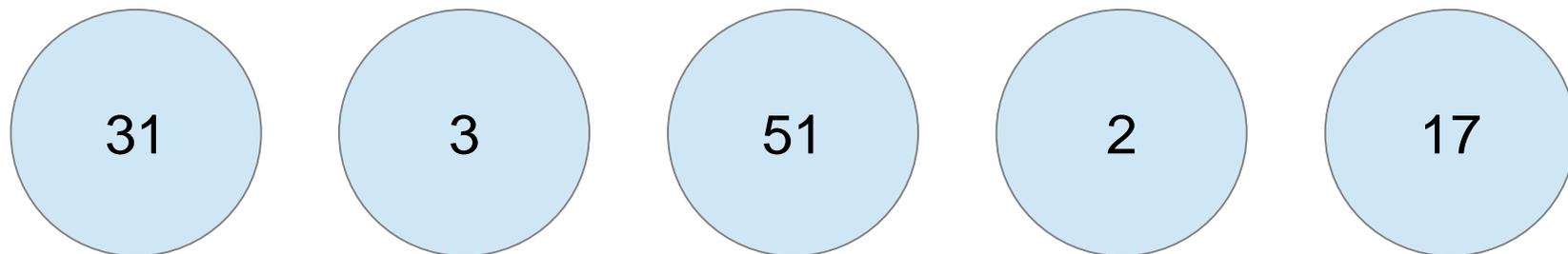


Cluster 5

Para cada cluster cria-se um cluster alternativo

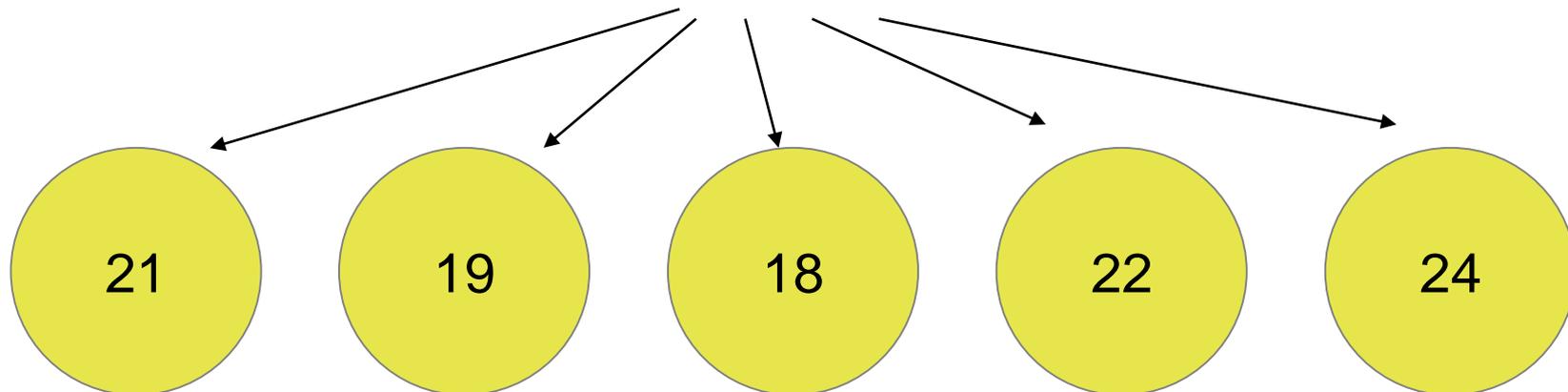


Clusters originais:



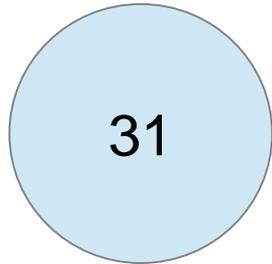
Cluster 1 Cluster 2 Cluster 3 Cluster 4 Cluster 5

Total = 104

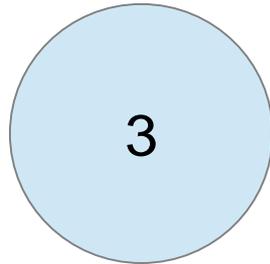


RCluster 1 RCluster 2 RCluster 3 RCluster 4 RCluster 5

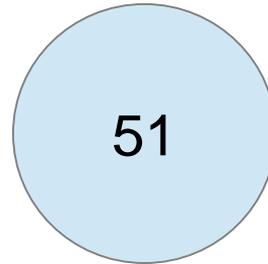
Clusters originais:



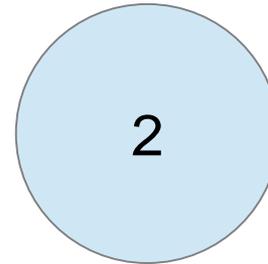
Cluster 1



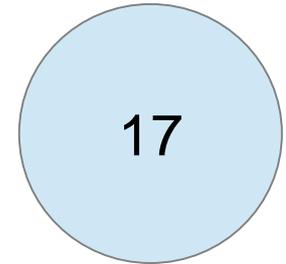
Cluster 2



Cluster 3

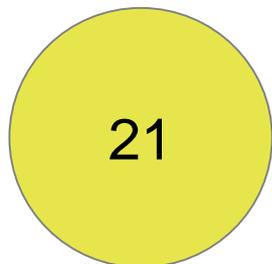


Cluster 4

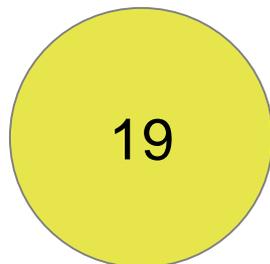


Cluster 5

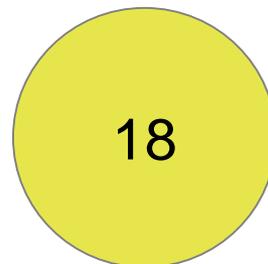
Obtemos a média μ e desvio padrão σ da distrib. normal



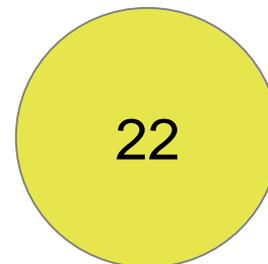
RCluster 1



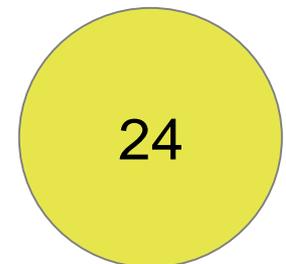
RCluster 2



RCluster 3

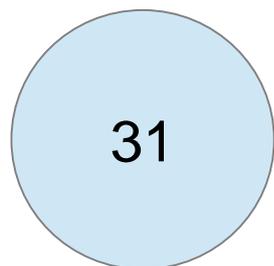


RCluster 4

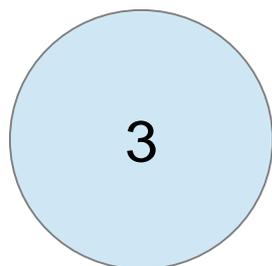


RCluster 5

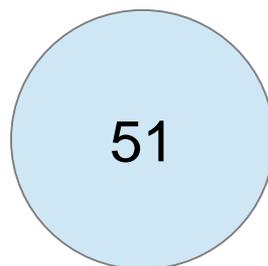
Clusters originais:



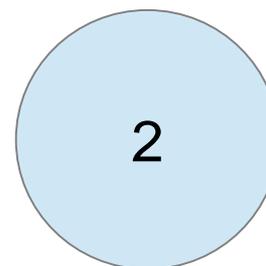
Cluster 1



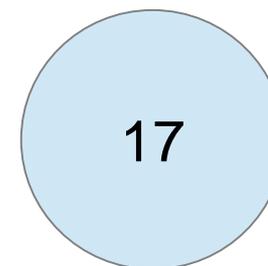
Cluster 2



Cluster 3

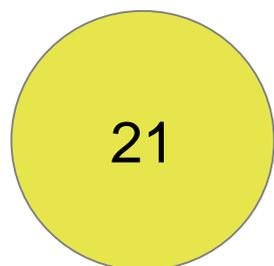


Cluster 4

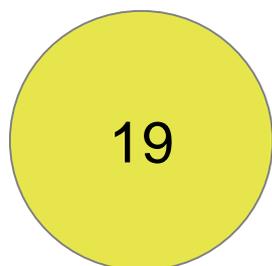


Cluster 5

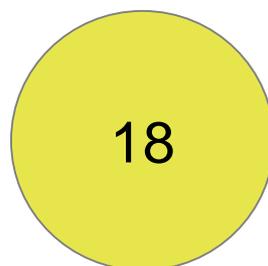
Excluimos todos aqueles em que o número de fotos do mesmo está a uma distância 2σ da média μ (~95%)



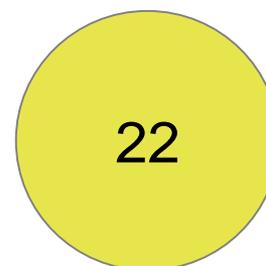
RCluster 1



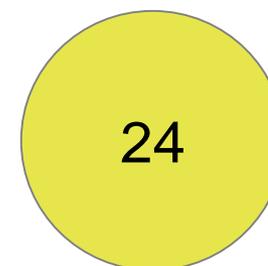
RCluster 2



RCluster 3

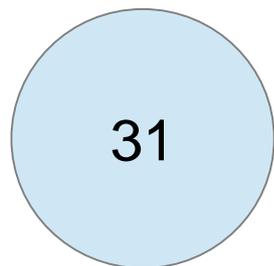


RCluster 4

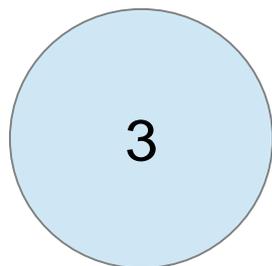


RCluster 5

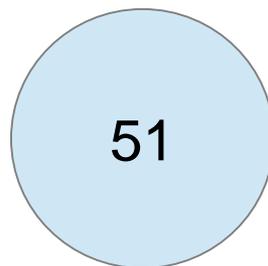
Clusters originais:



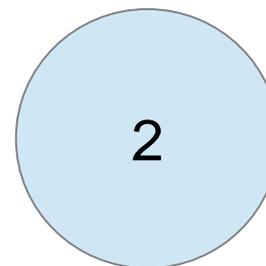
Cluster 1



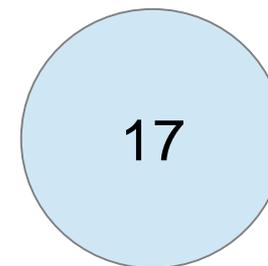
Cluster 2



Cluster 3

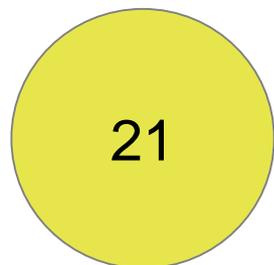


Cluster 4

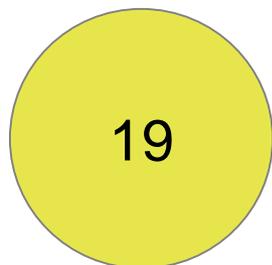


Cluster 5

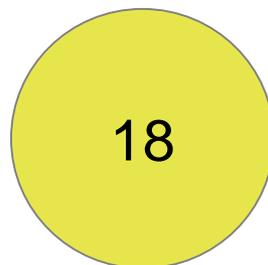
A ideia é excluir clusters que provavelmente não refletem um PDI da cidade



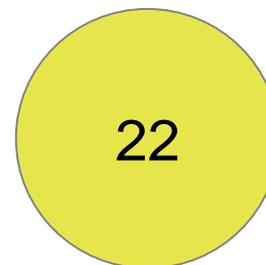
RCluster 1



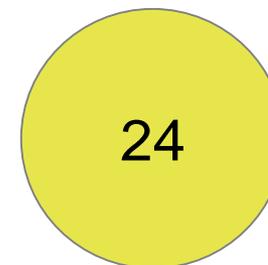
RCluster 2



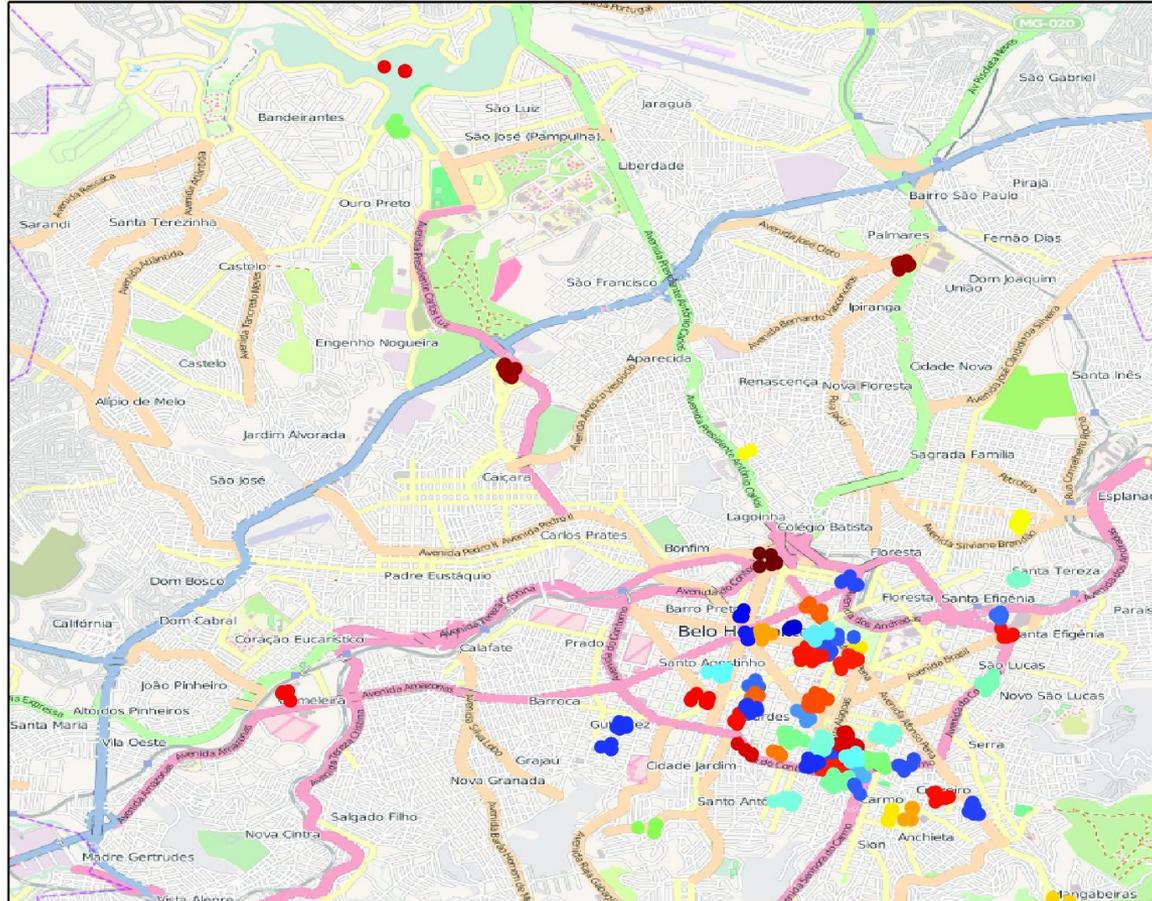
RCluster 3



RCluster 4



RCluster 5



Pontos de interesse

Padrões Sociais, Econômicos e Culturais



Estudo de diferenças culturais

You are What you Eat (and Drink): Identifying Cultural Boundaries by Analyzing Food & Drink Habits in Foursquare.
ICWS'15



Grande desafio: encontrar dados apropriados para uso

- **Métodos tradicionais:** Questionários
 - Não escalam
 - Difícil de detectar mudanças dinâmicas

Grande desafio: encontrar dados apropriados para uso

- **Métodos tradicionais:** Questionários
 - Não escalam
 - Difícil de detectar mudanças dinâmicas

É possível propor algum método alternativo?

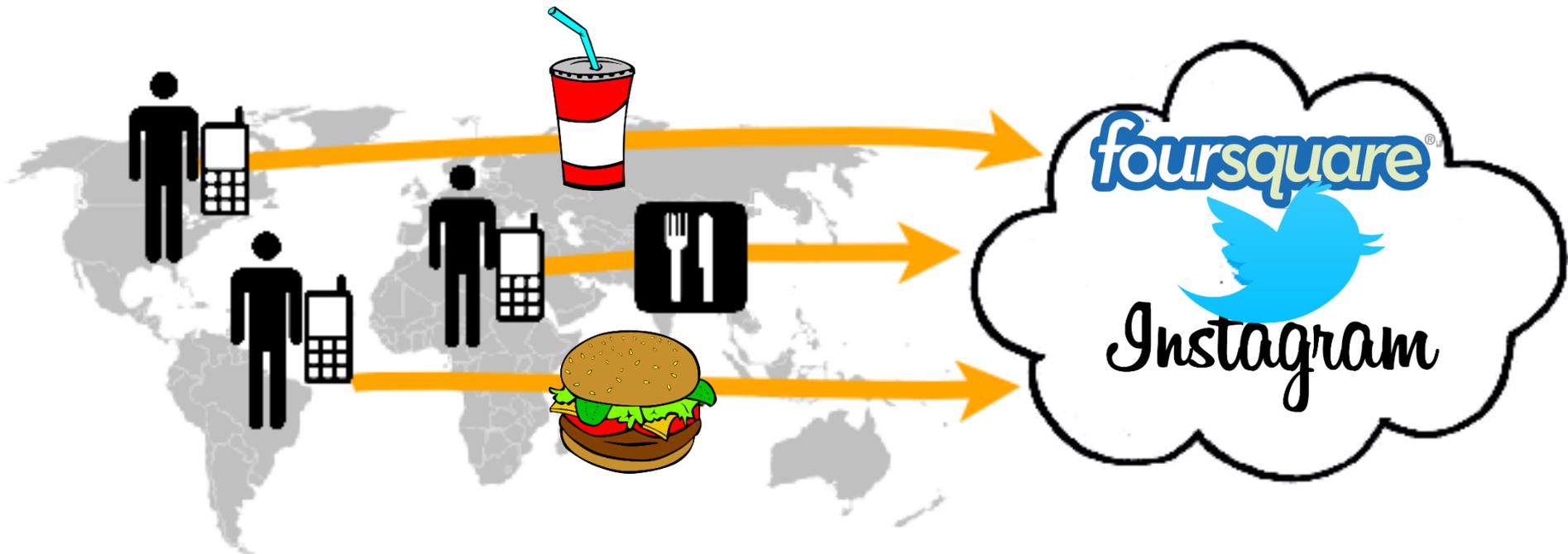
Você é o que você come



Hábitos alimentares e de bebida são elementos fundamentais em uma cultura

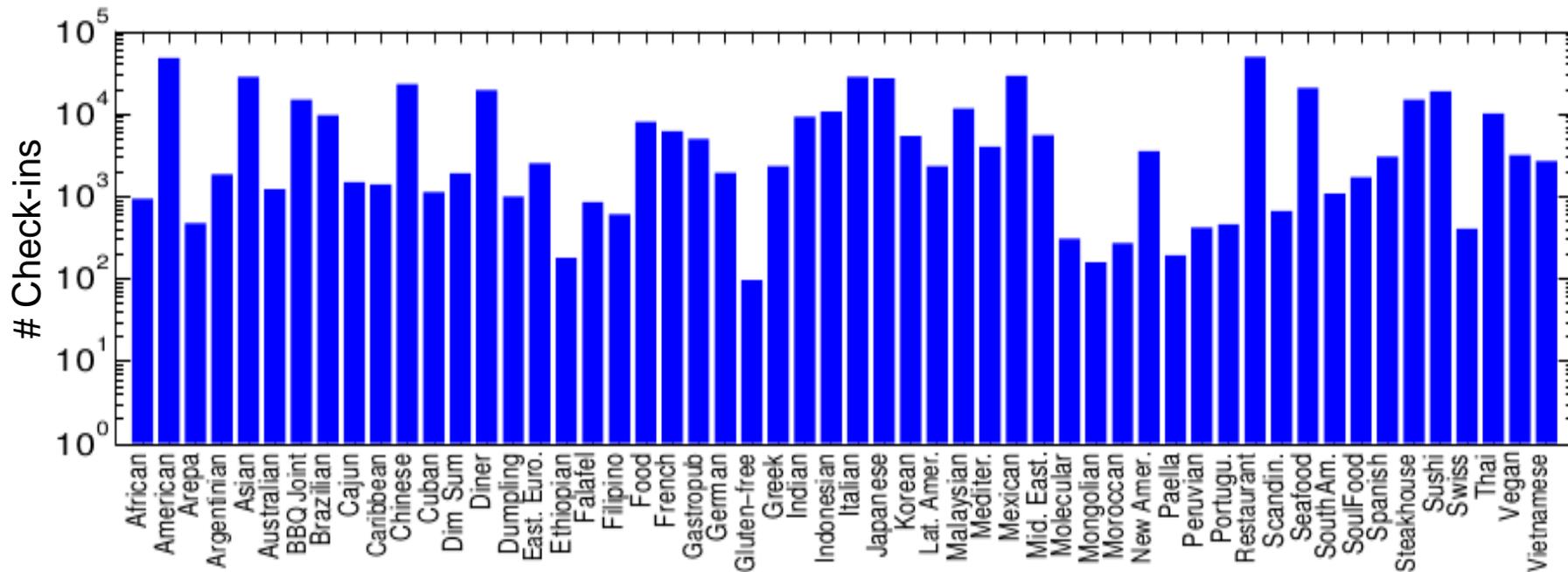


Sensoriamento de atividades humanas em larga escala!

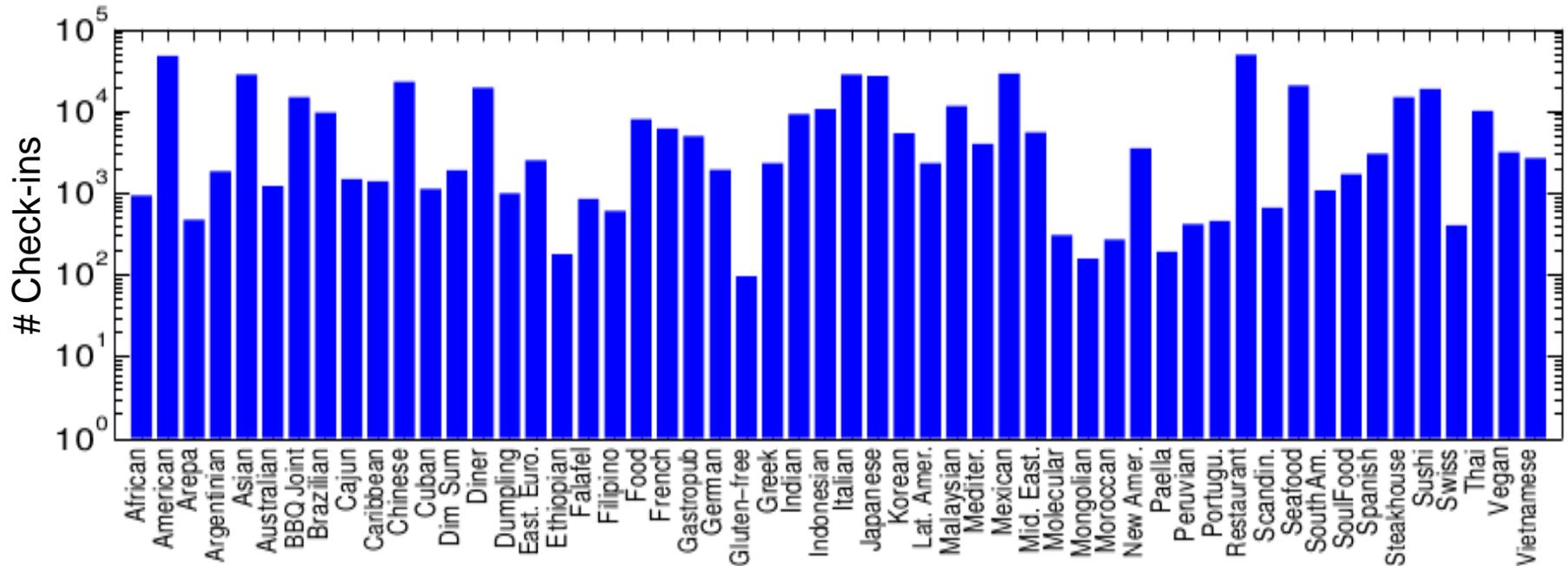


Oportunidade sem precedentes para estudar diferenças culturais em escala global e baixo custo

Categorias Slow Food

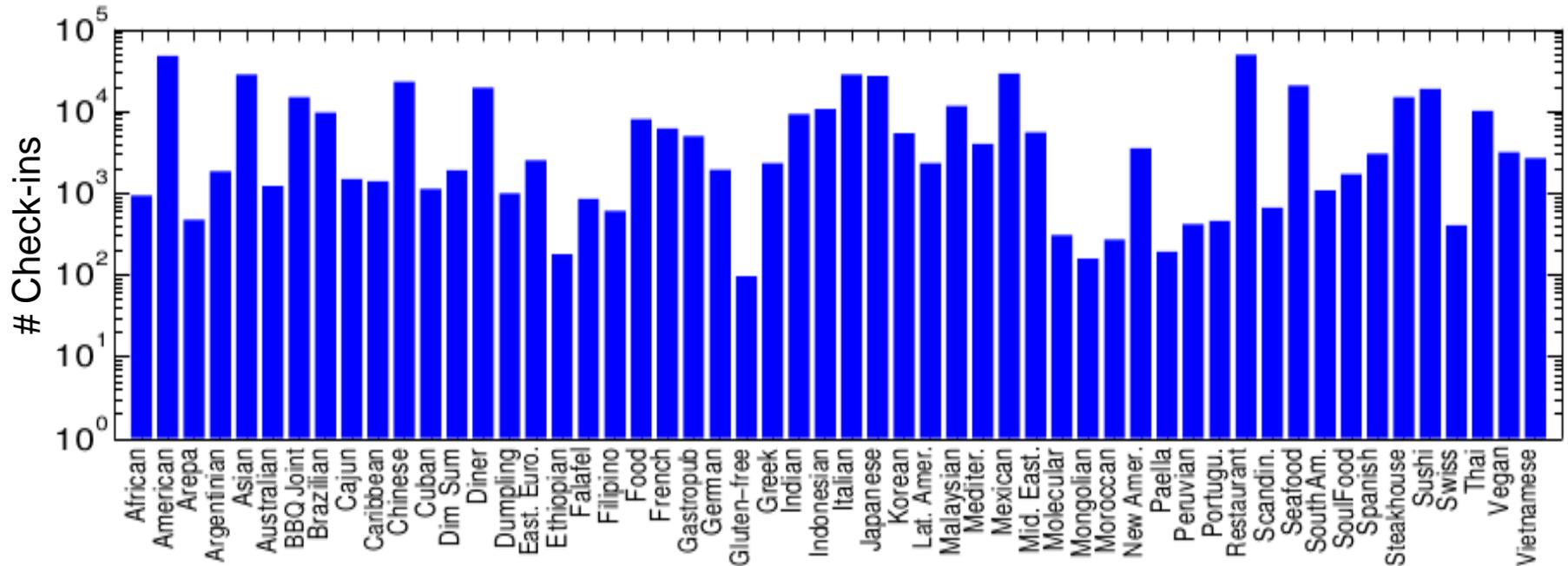


Mapeando os hábitos



Mapeia cada usuário n_i em $F_i = f_{1i}, f_{2i}, \dots, f_{mi}$

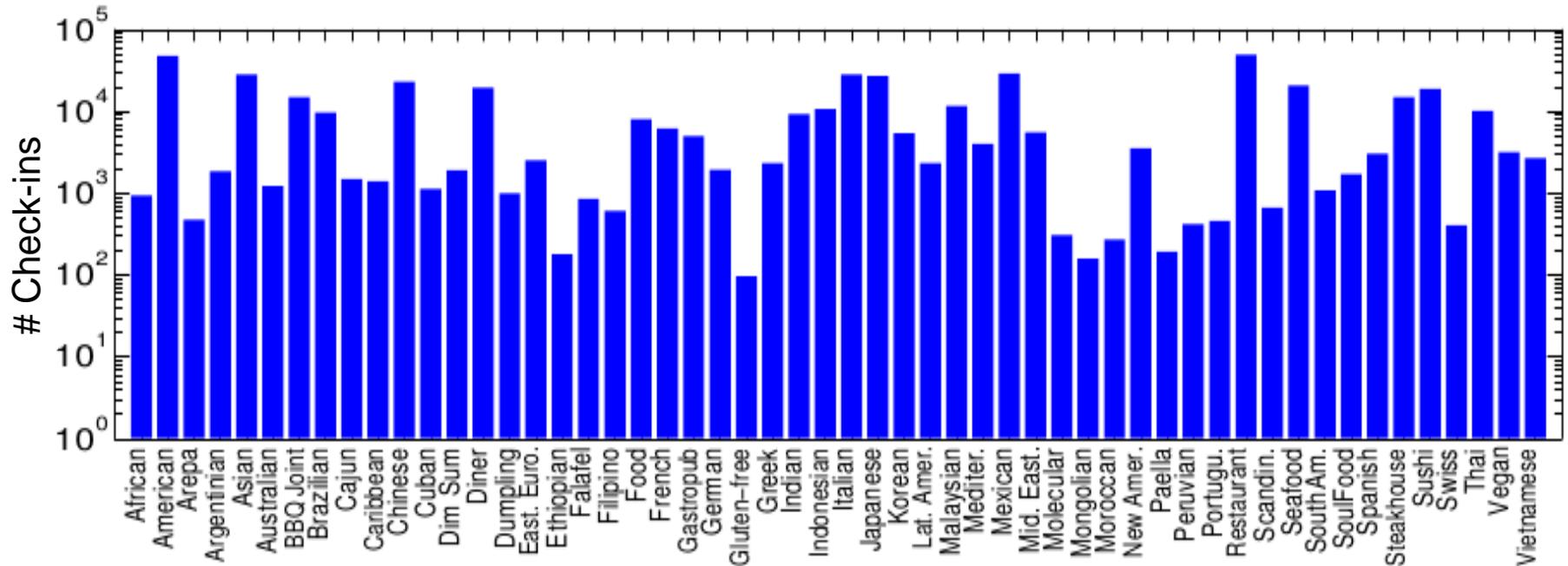
Mapeando os hábitos



Mapeia cada usuário n_i em $F_i = f_{1i}, f_{2i}, \dots, f_{mi}$

Como perguntas em um survey!

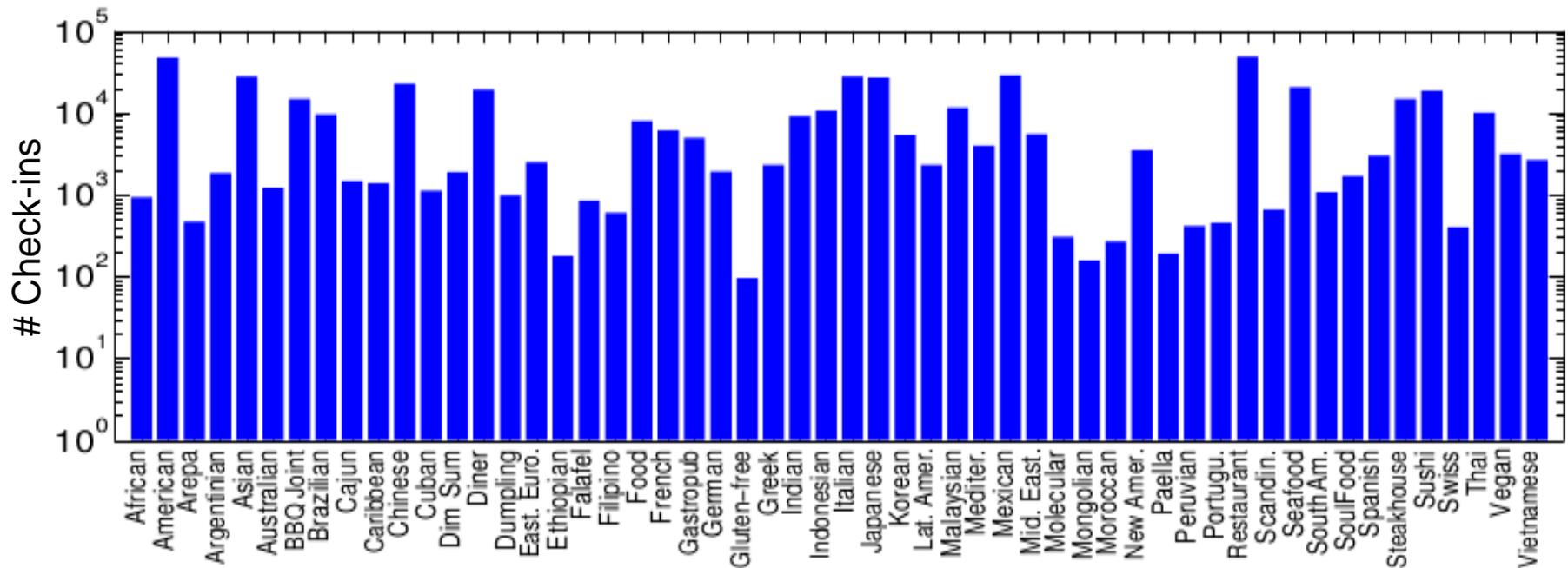
Mapeando os hábitos



Mapeia cada usuário n_i em $F_i = f_{1i}, f_{2i}, \dots, f_{m_i}$

$f_{k^i} = 0|1$ representa se o usuário n_i gosta de f_k

Mapeando os hábitos



Mapeia cada usuário n_i em $F_i = f_{1i}, f_{2i}, \dots, f_{m_i}$

$f_{k^i} = 0|1$ representa se o usuário n_i gosta de f_k

Respostas dos usuários

Análise cultural de indivíduos





Criamos um grafo em que:

Nós são usuários





Uma aresta existe se a similaridade dos usuários é maior do que s



Análise cultural de indivíduos

50% de similaridade
Entre 2
vetores de preferência



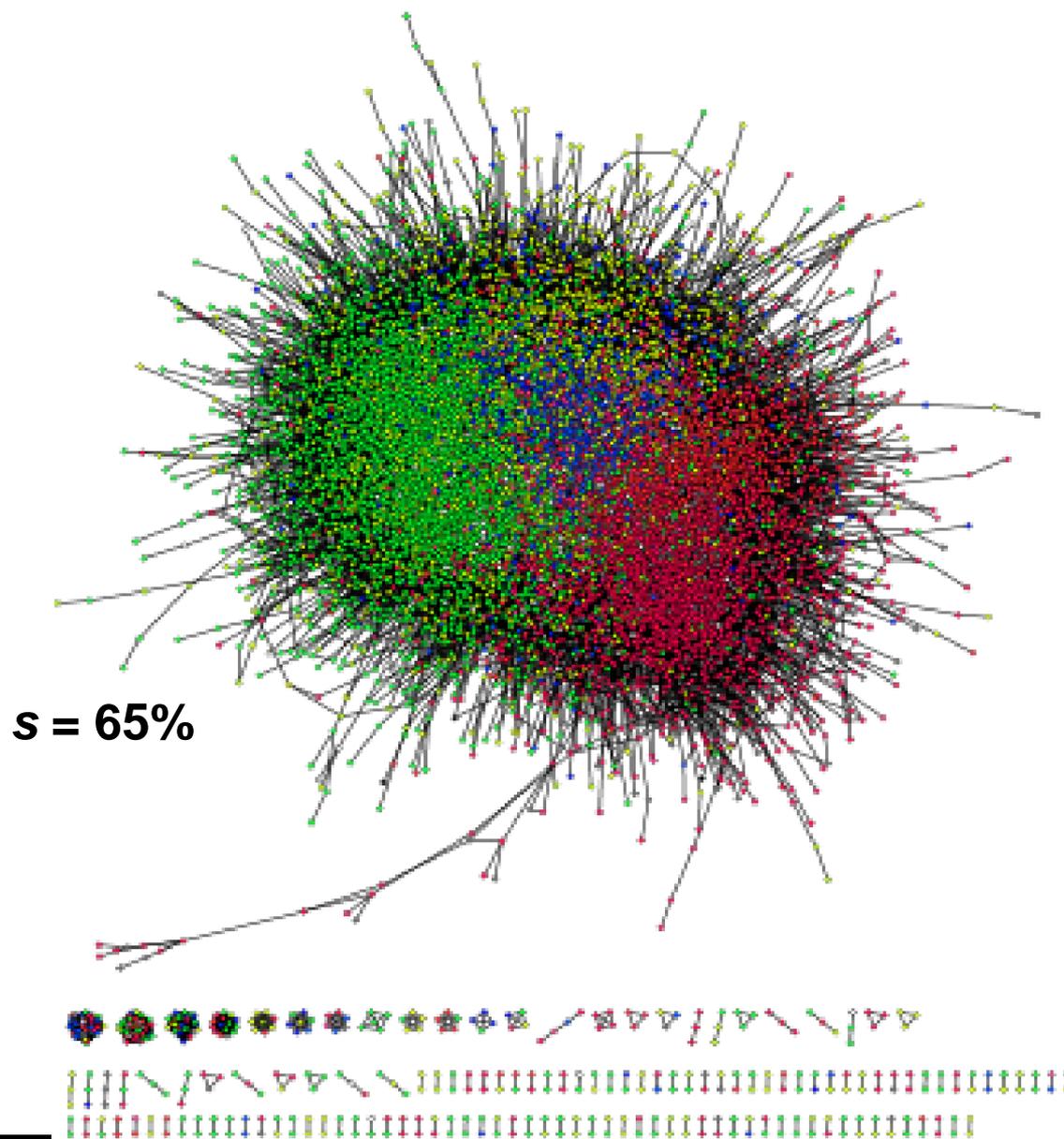
Análise cultural de indivíduos

90% de similaridade

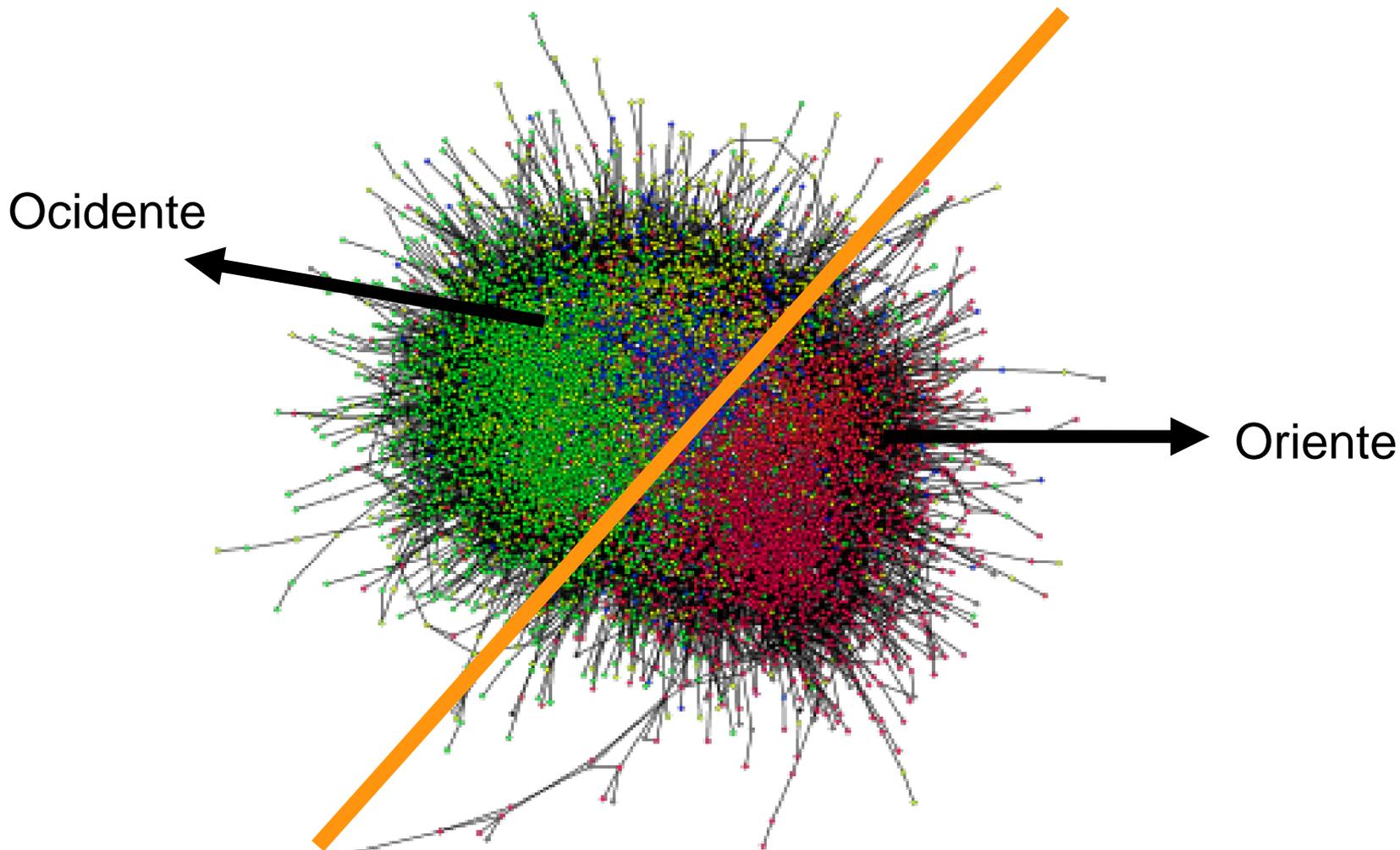


A cor do nó indica a região do usuário:

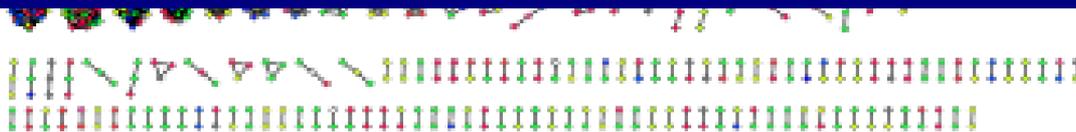
- África
- Ásia
- Europa
- America do Norte
- América do Sul
- Oceania



Force-direct
layout algorithm

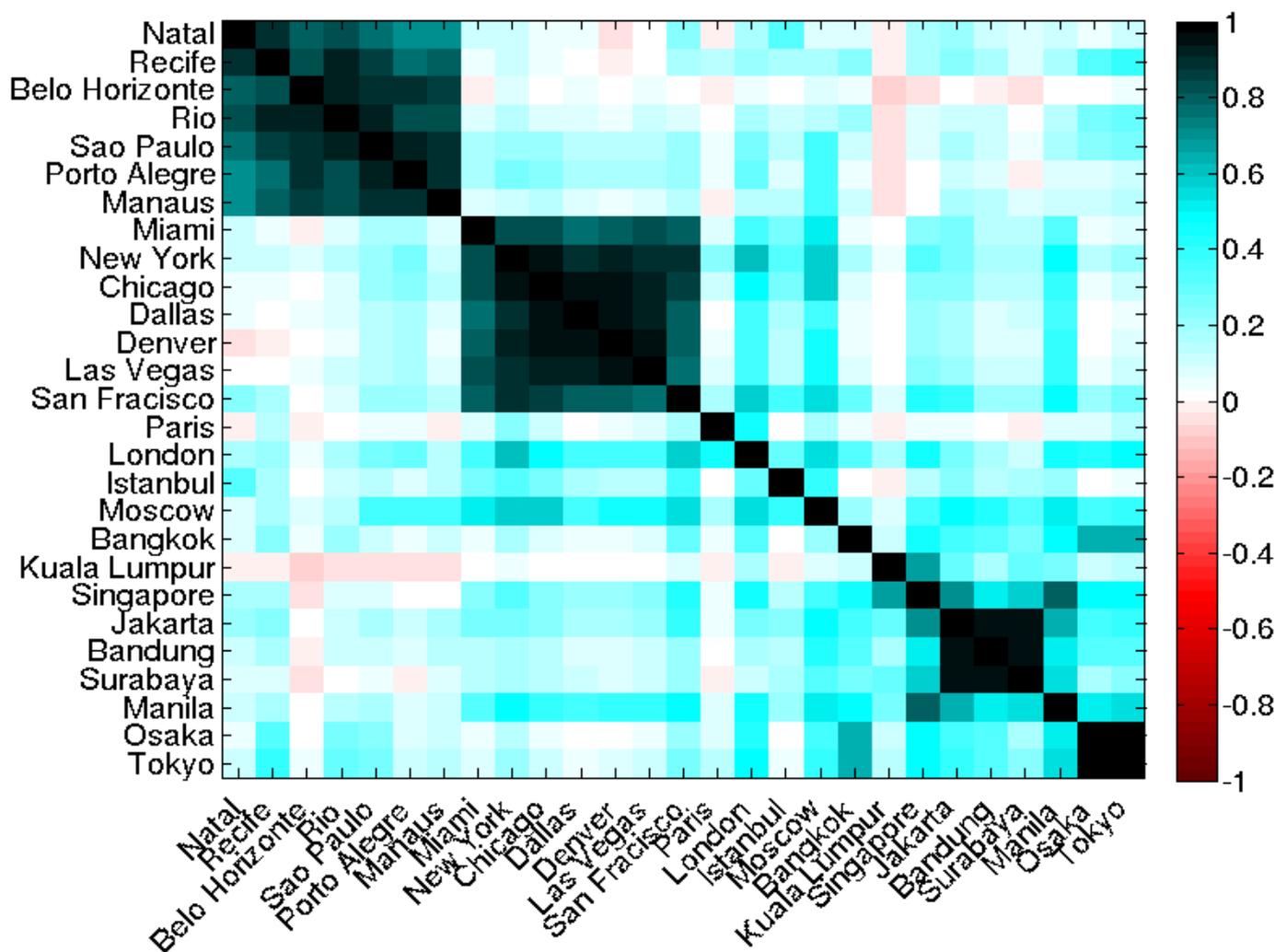


Pessoas do ocidente tendem a ser mais similares entre si



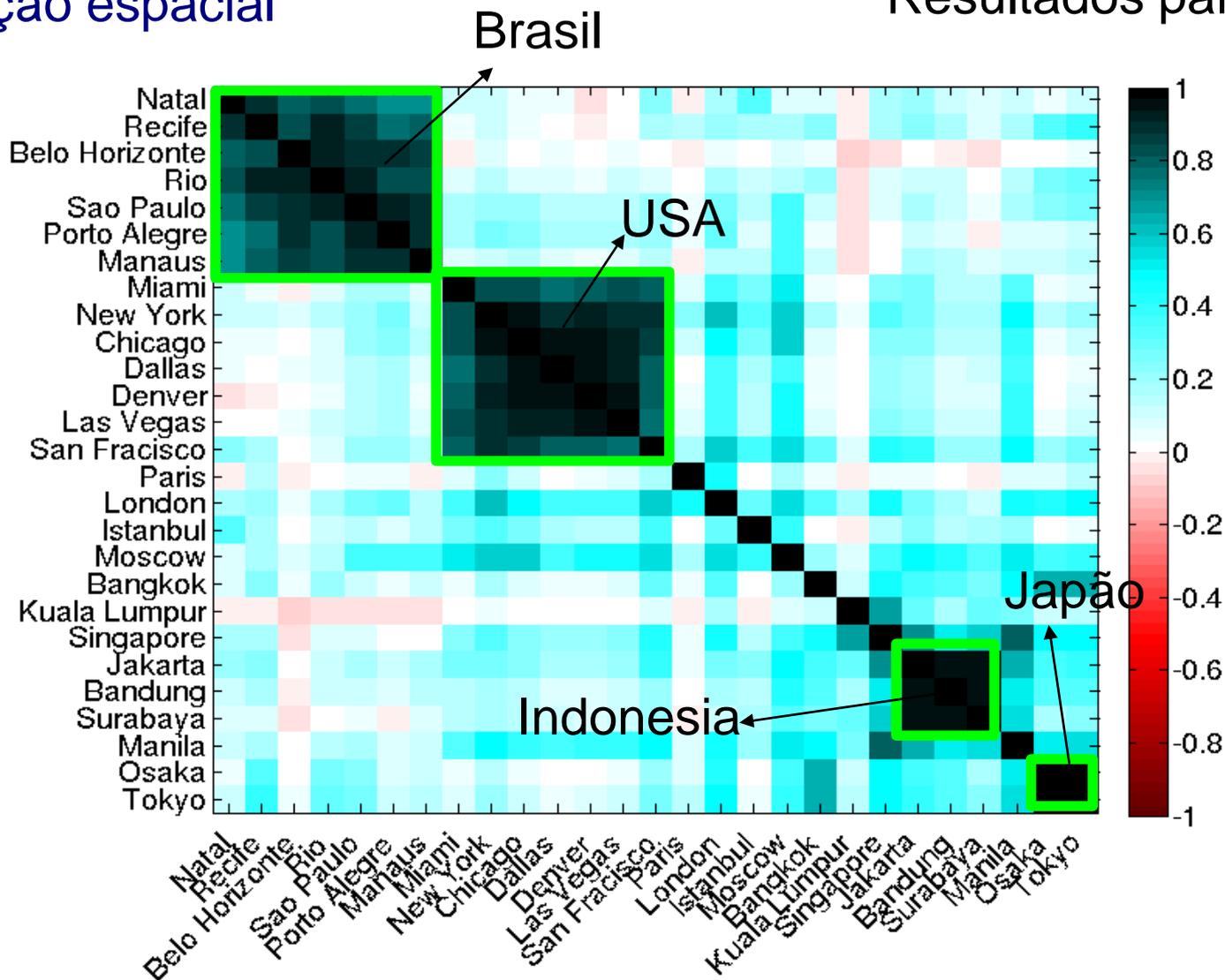
Avaliação espacial

Resultados para cidades



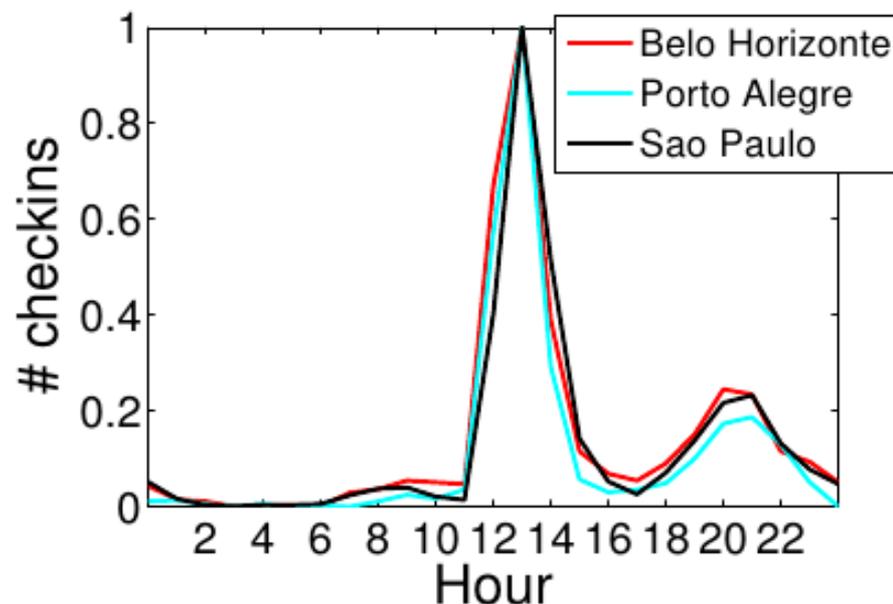
Avaliação espacial

Resultados para cidades

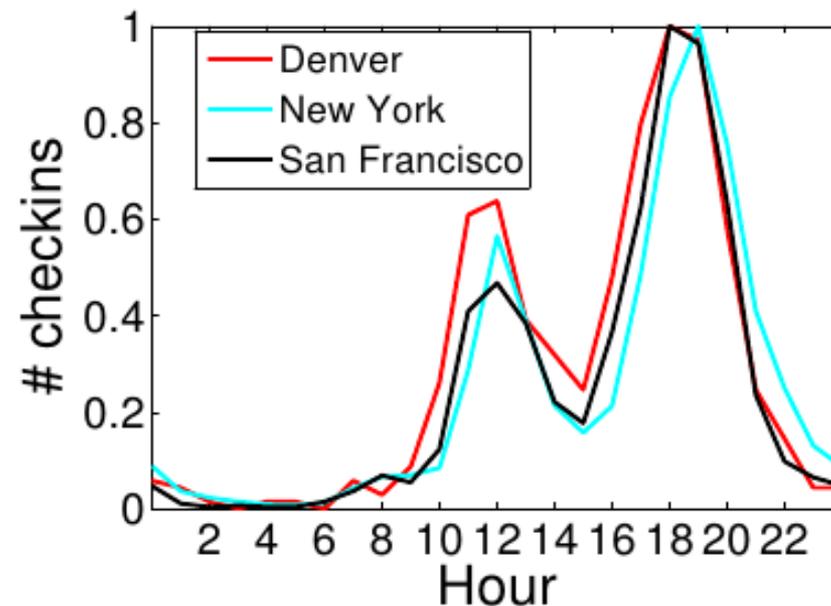


Avaliação temporal

Dias de semana



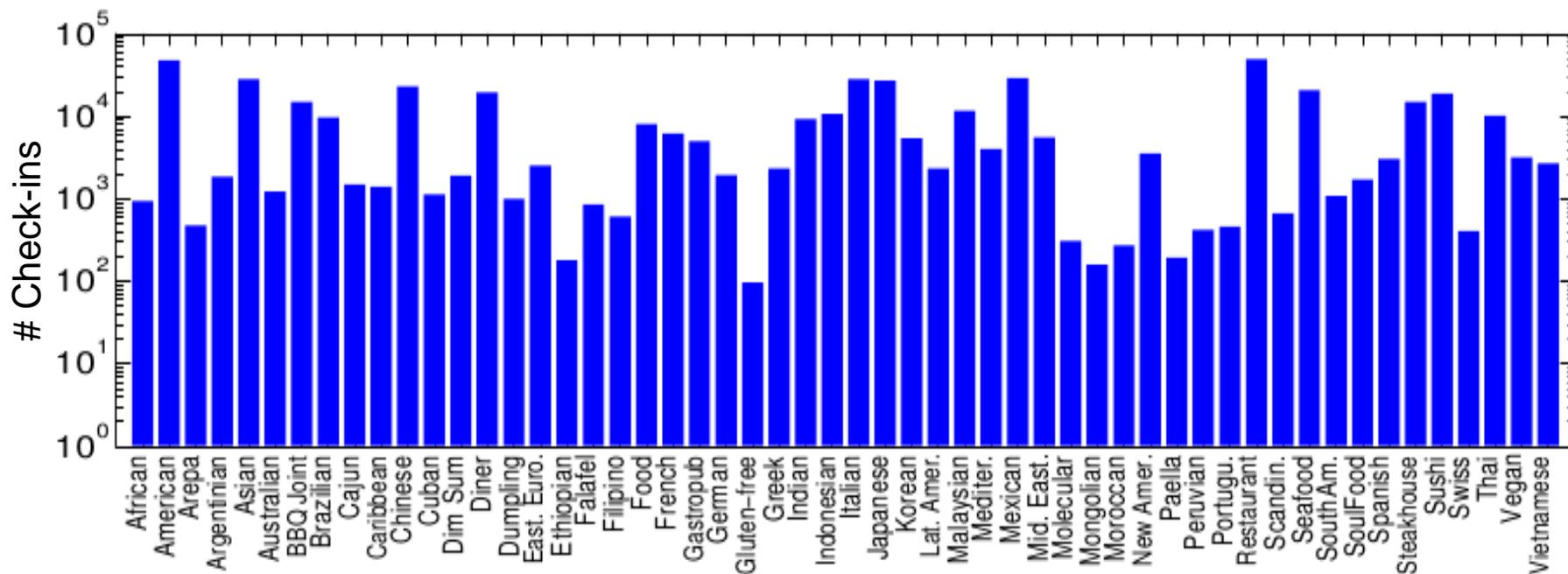
Cidades brasileiras



Cidades dos EUA

A maioria das cidades seguem o padrão geral do país

Nossa abordagem:



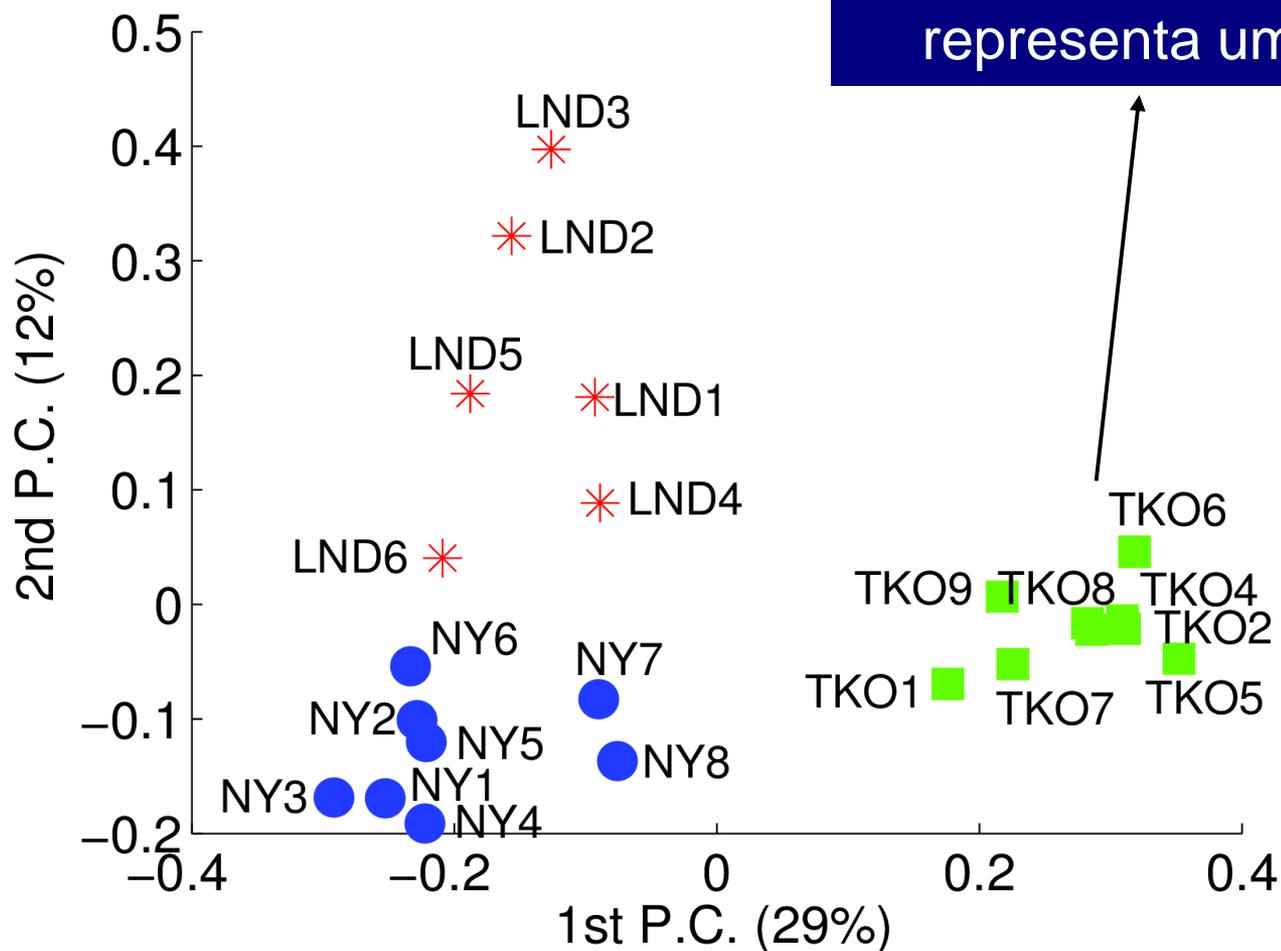
Vetor de preferências para a área
(tempo e espaço)



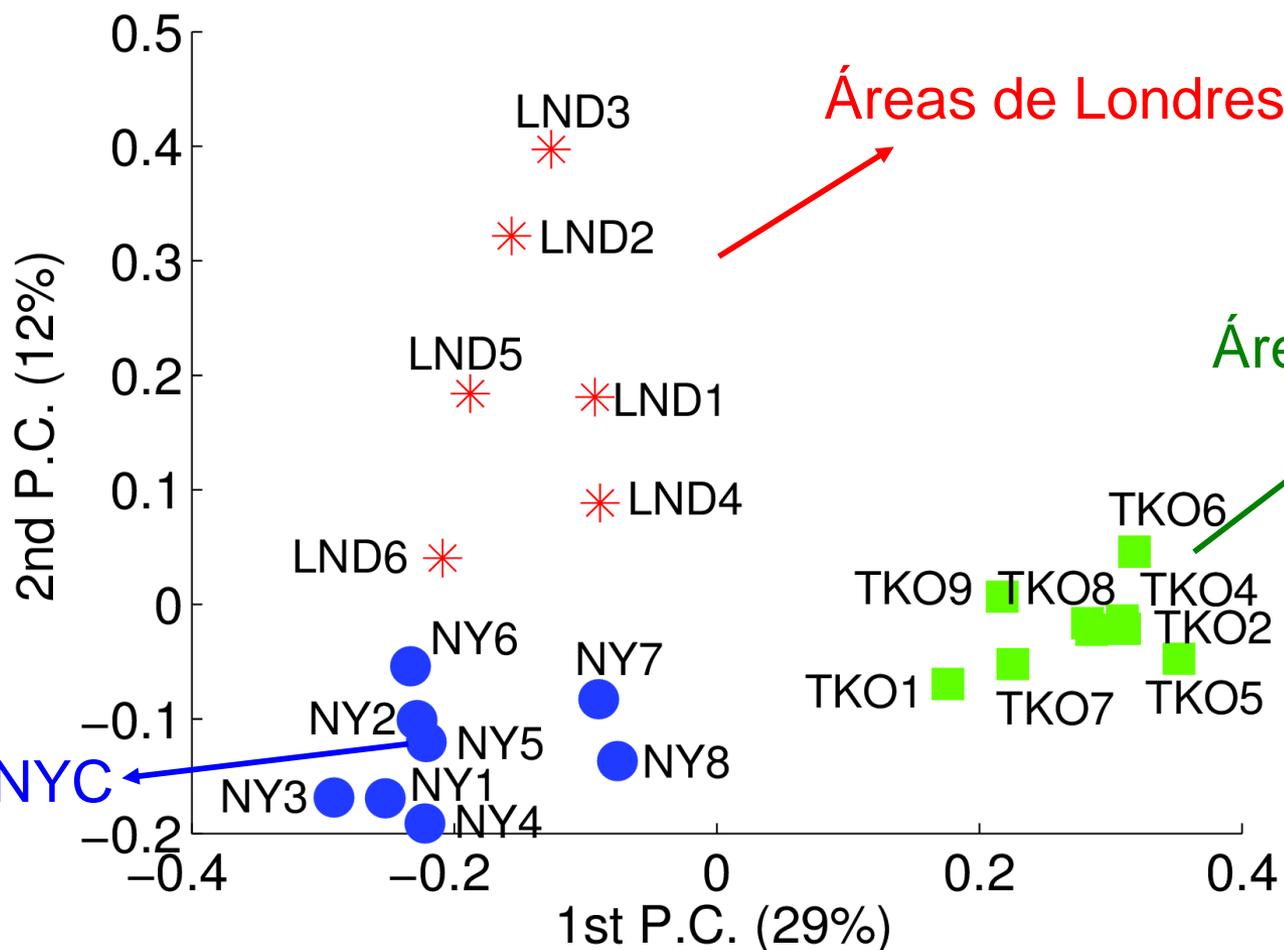
Principal Component Analysis (PCA)

k-means para agrupar áreas no espaço definido pelos PCs

Agrupando áreas dentro da cidade

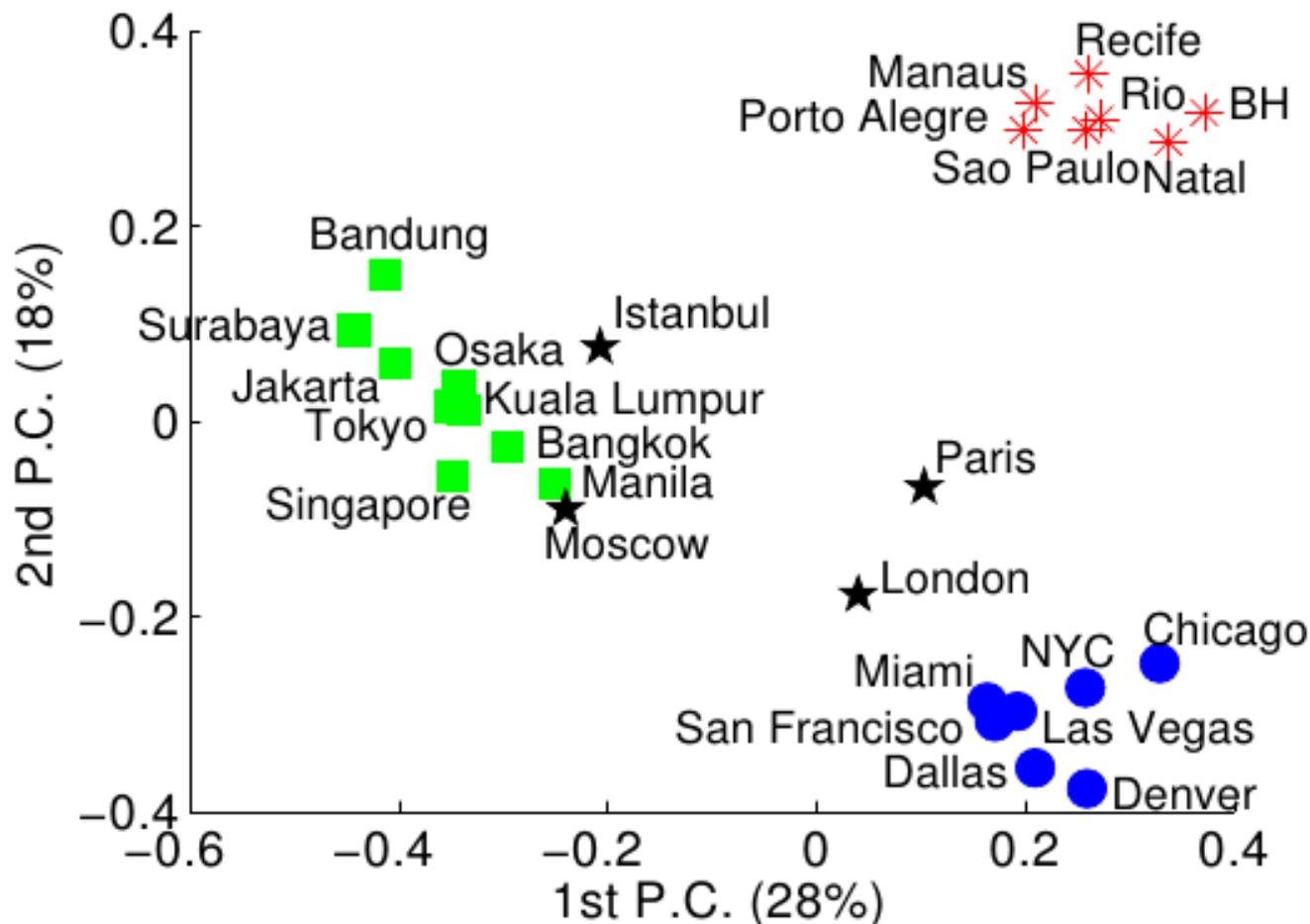


Agrupando áreas dentro da cidade



k = 3 (3 cidades diferentes)

Agrupando cidades

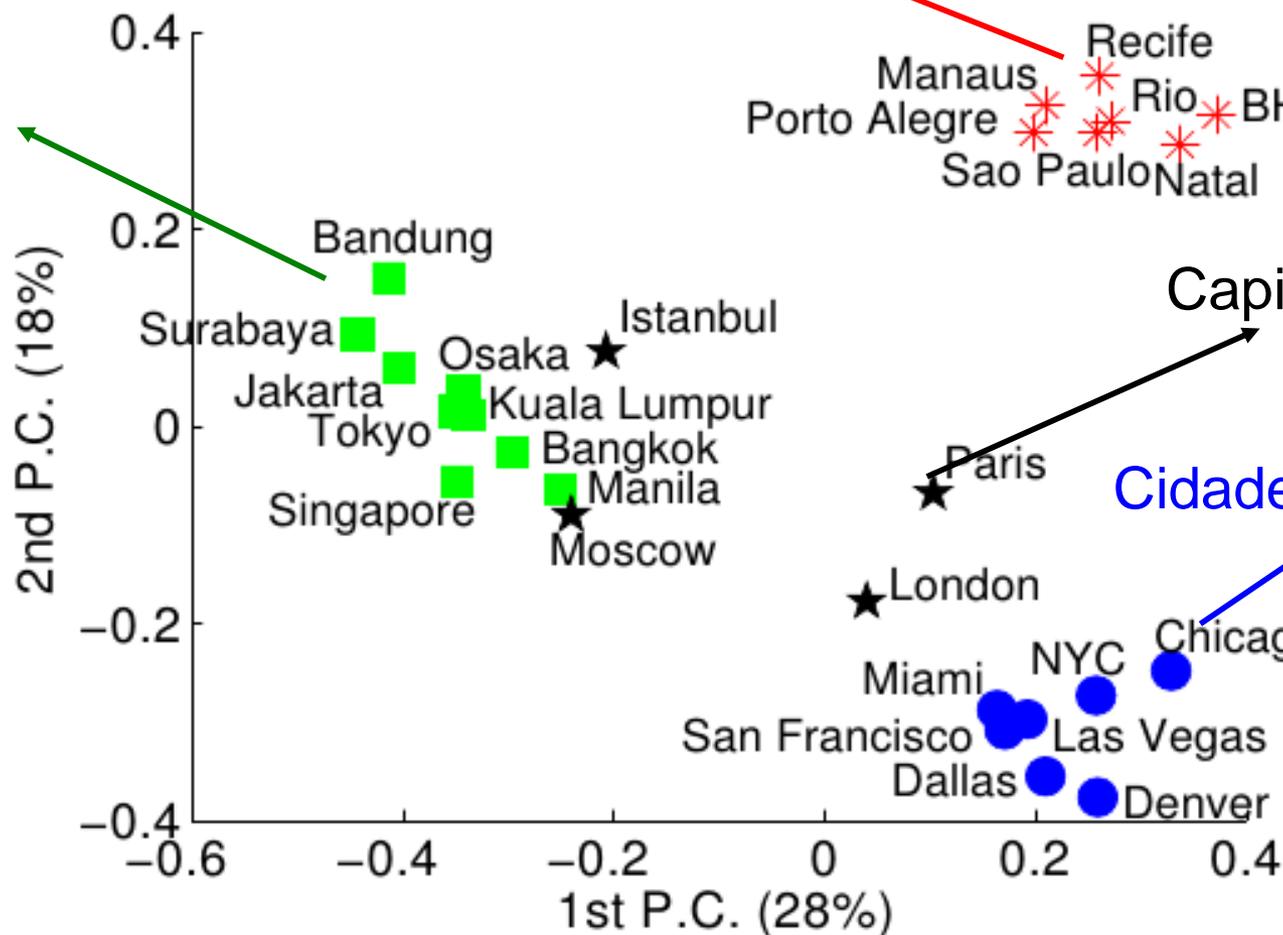


$k = 4$ (4 regiões distintas)

Agrupando cidades

Cidades brasileiras

Cidades asiáticas

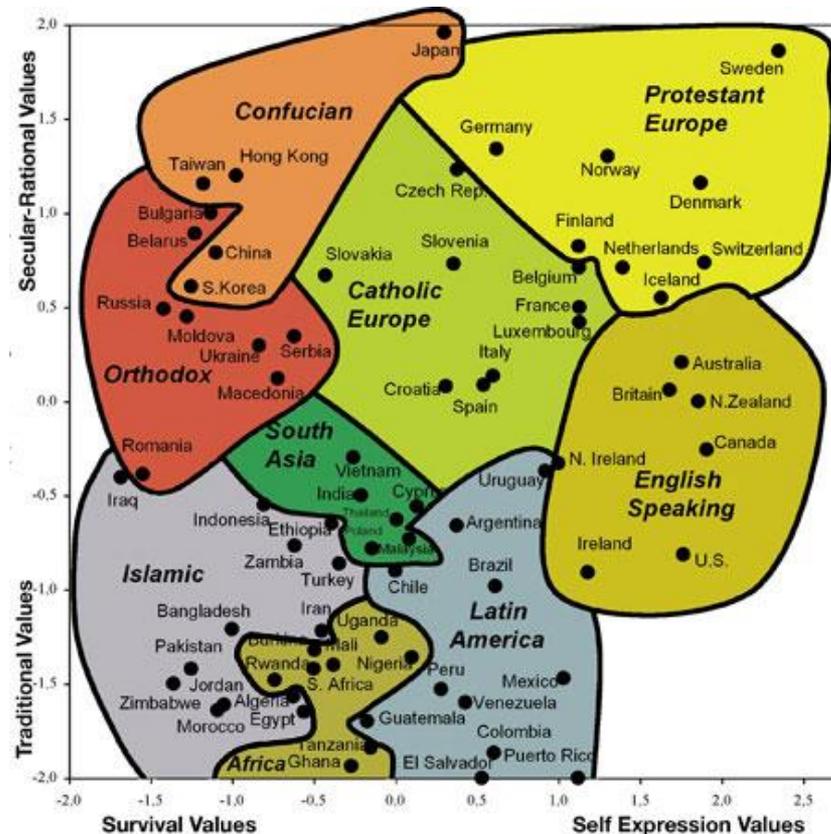


Capitais europeias

Cidades americanas

World Value Survey

Um estudo abrangente sobre diferenças culturais



World Value Survey

Um estudo abrangente sobre diferenças culturais

Os resultados são praticamente os mesmos!

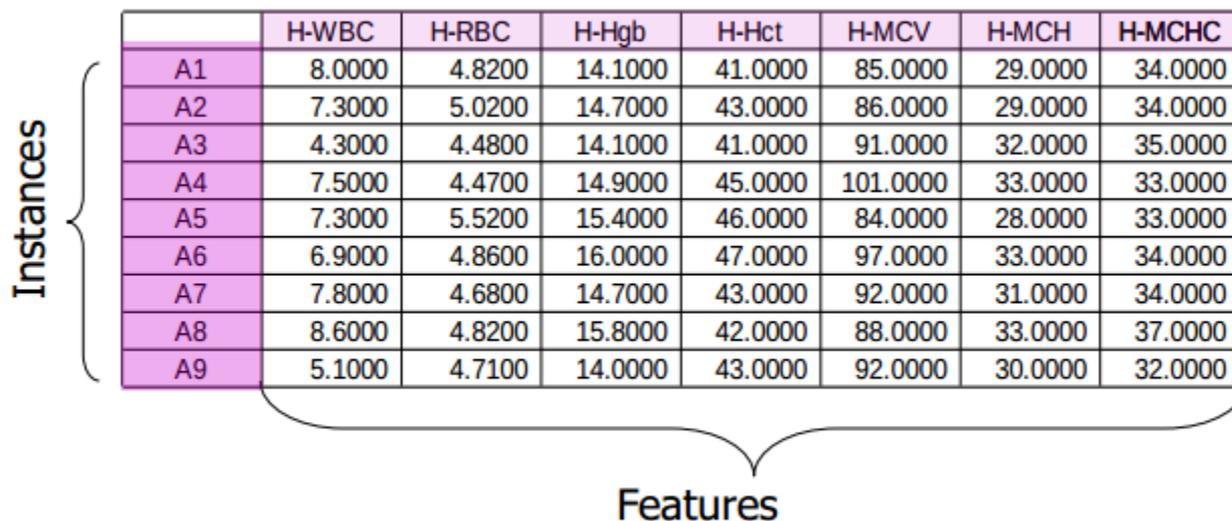
Mesmo resultado usando **dados com 2 anos de diferença**

Redução da dimensionalidade: representa dados em menos dimensões

- Aprendizado mais fácil – menos parâmetros
- Visualização – mostrar muitas dimensões em 2D
- Descobrir "dimensionalidade intrínseca" dos dados
 - Redução de ruído

Redução da dimensionalidade

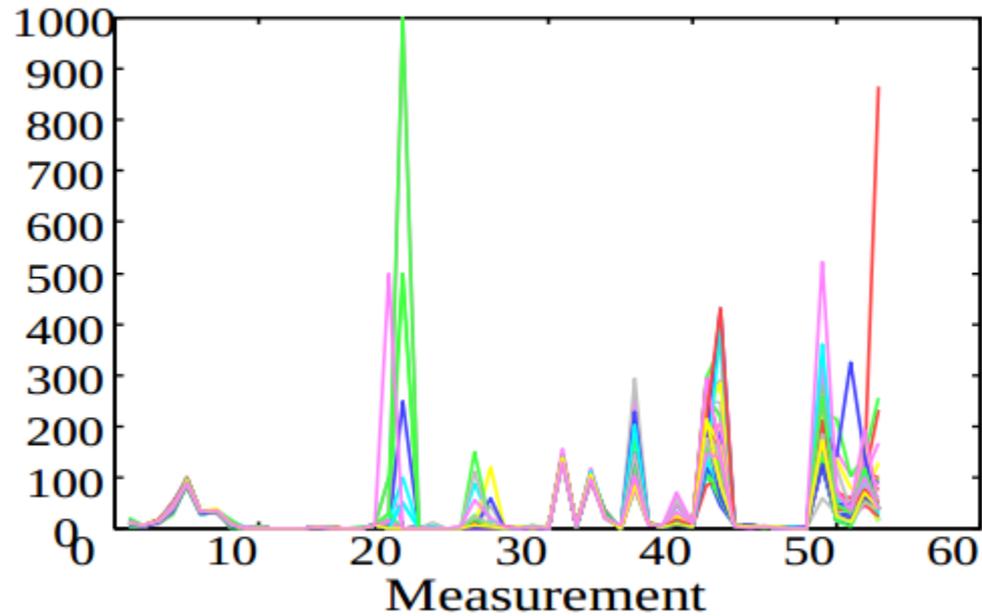
- Matrix format (65x53)



	H-WBC	H-RBC	H-Hgb	H-Hct	H-MCV	H-MCH	H-MCHC
A1	8.0000	4.8200	14.1000	41.0000	85.0000	29.0000	34.0000
A2	7.3000	5.0200	14.7000	43.0000	86.0000	29.0000	34.0000
A3	4.3000	4.4800	14.1000	41.0000	91.0000	32.0000	35.0000
A4	7.5000	4.4700	14.9000	45.0000	101.0000	33.0000	33.0000
A5	7.3000	5.5200	15.4000	46.0000	84.0000	28.0000	33.0000
A6	6.9000	4.8600	16.0000	47.0000	97.0000	33.0000	34.0000
A7	7.8000	4.6800	14.7000	43.0000	92.0000	31.0000	34.0000
A8	8.6000	4.8200	15.8000	42.0000	88.0000	33.0000	37.0000
A9	5.1000	4.7100	14.0000	43.0000	92.0000	30.0000	32.0000

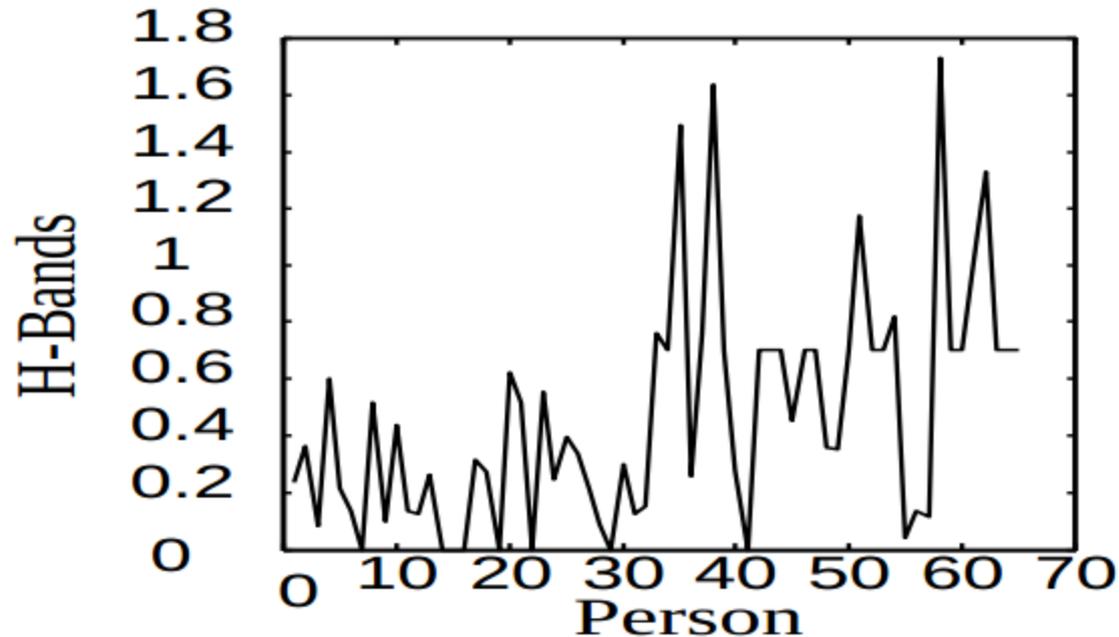
Difficult to see the correlations between the features...

- Spectral format (65 curves, one for each person)



Difficult to compare the different patients...

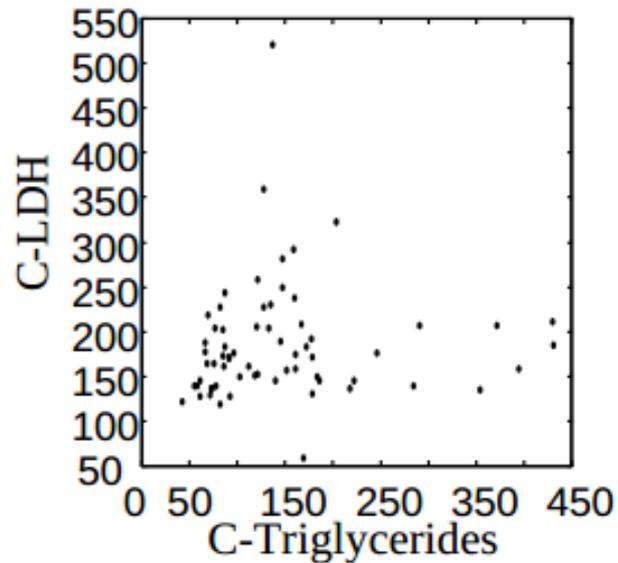
- Spectral format (53 pictures, one for each feature)



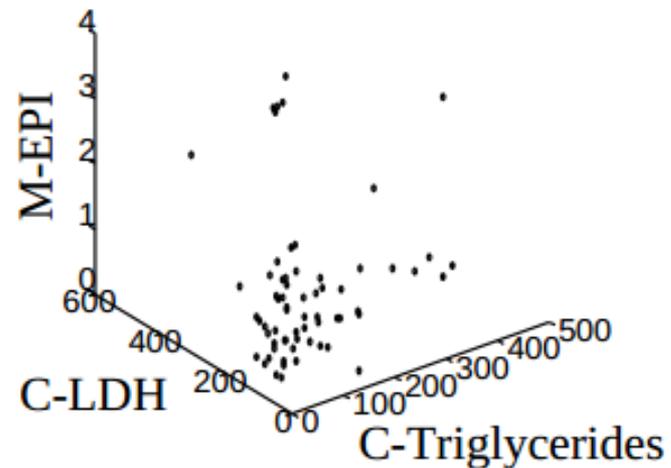
Difficult to see the correlations between the features...

8

Bi-variate



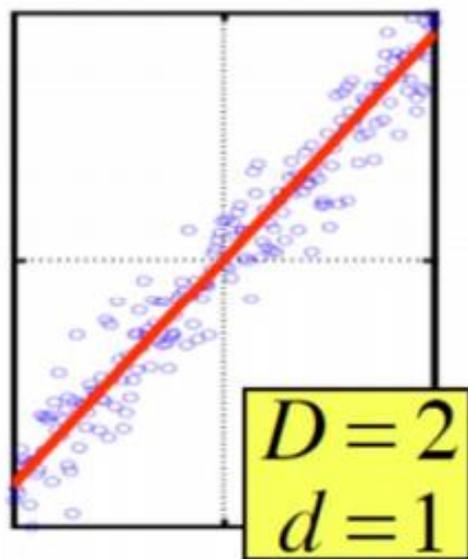
Tri-variate



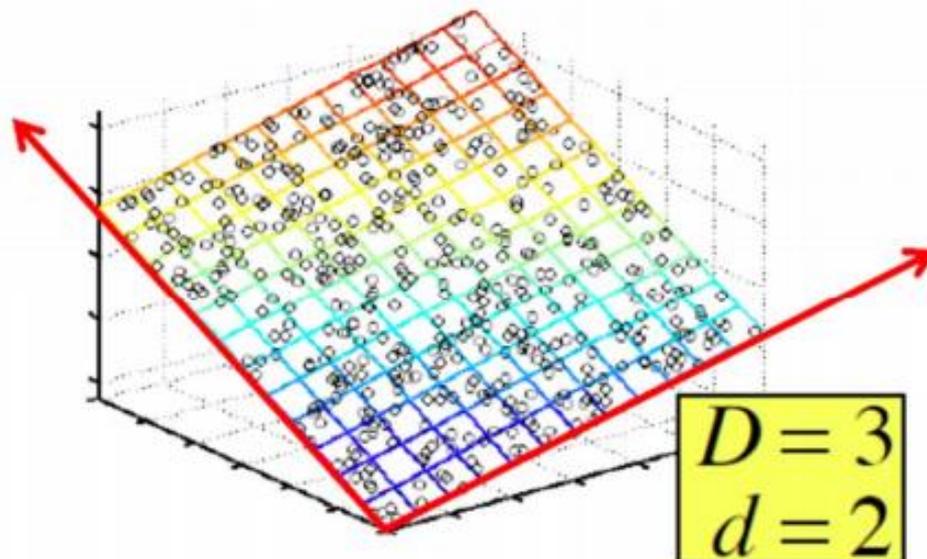
How can we visualize the other variables???

... difficult to see in 4 or higher dimensional spaces...

Hipótese: dados encontram-se aproximadamente em um espaço dimensional menor

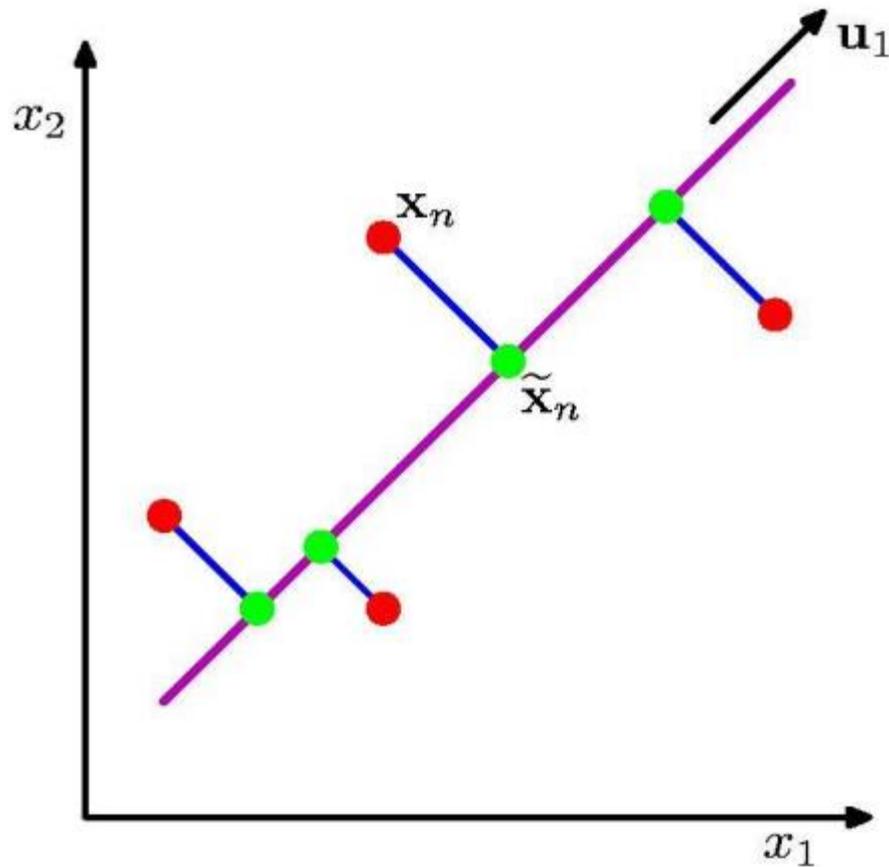


Redução 2 pra 1 dimensão



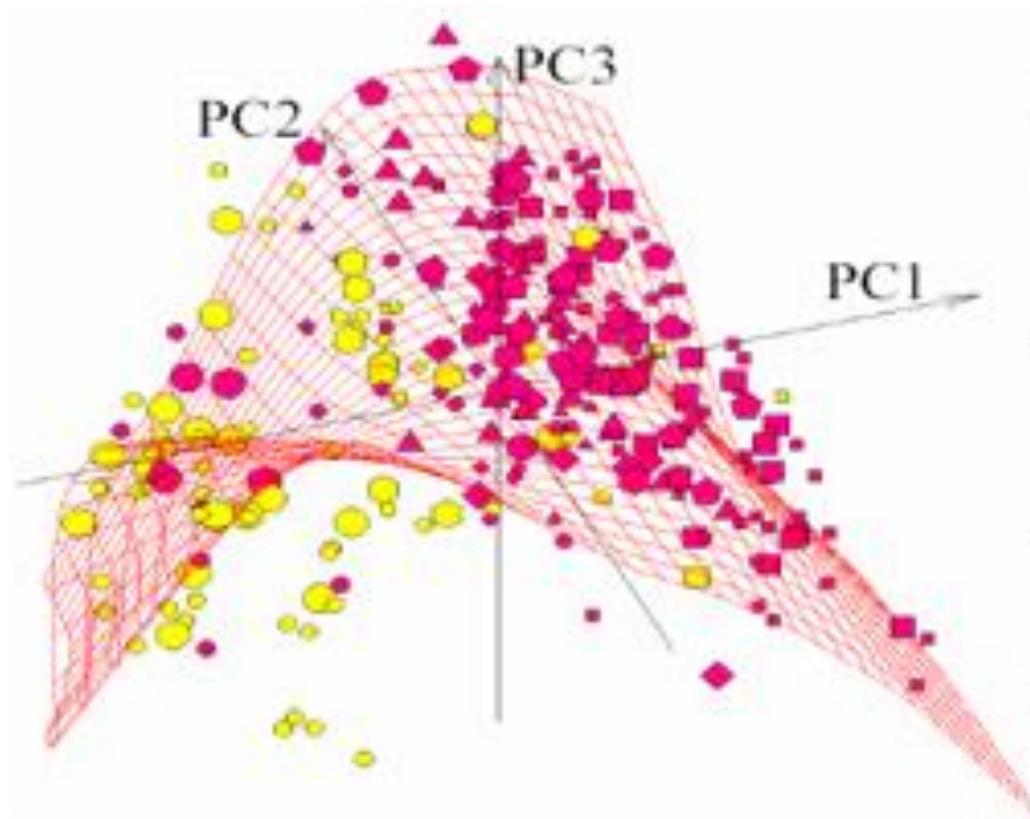
Redução 3 pra 2 dimensão

Imagem de Yi Zhang



Projeção ortogonal dos dados em um espaço dimensional menor:

- maximizar a variação dos dados projetados (linha roxa)
- minimizar a distância entre os pontos de dados e as projeções (linhas azuis)

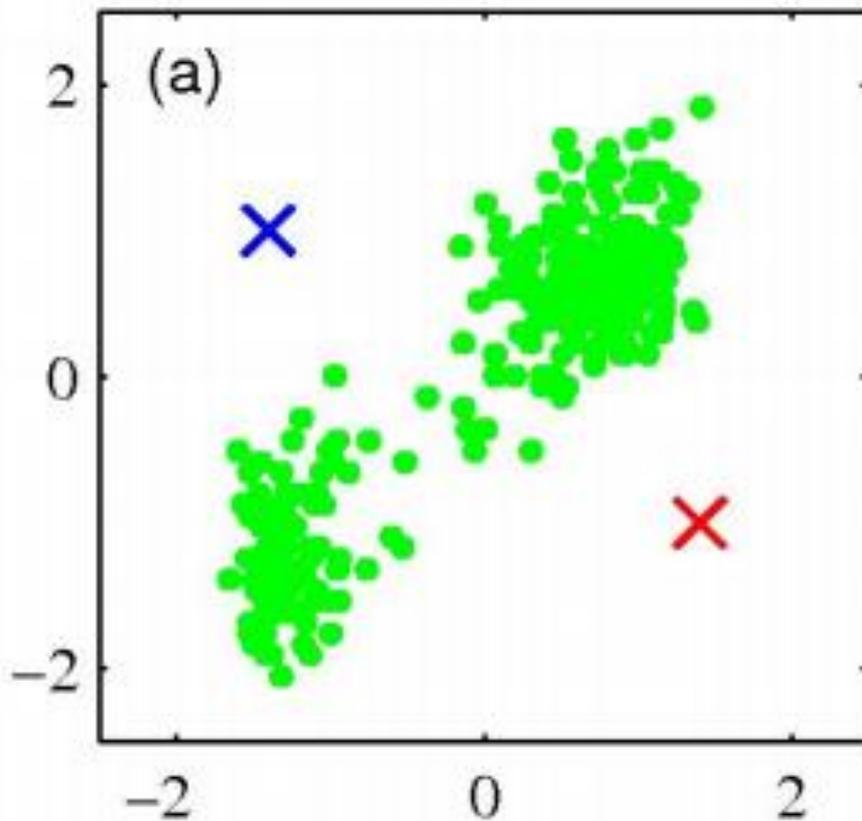


K-means

Algoritmo de agrupamento baseado em particionamento

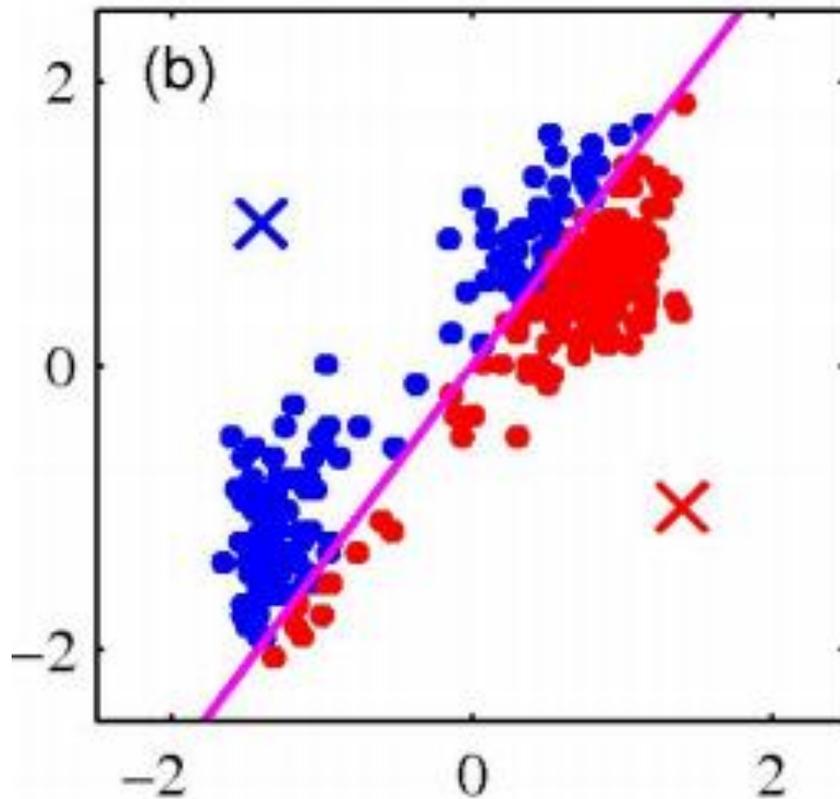
- **Inicialização:** Escolher k pontos aleatórios para serem o centro do cluster
- **Alternar:**
 1. Atribuir os pontos para o centro mais próximo
 2. Mudar o centro do cluster para a média dos pontos atribuídos
- **Parar quando não mudanças significativas**

K-means



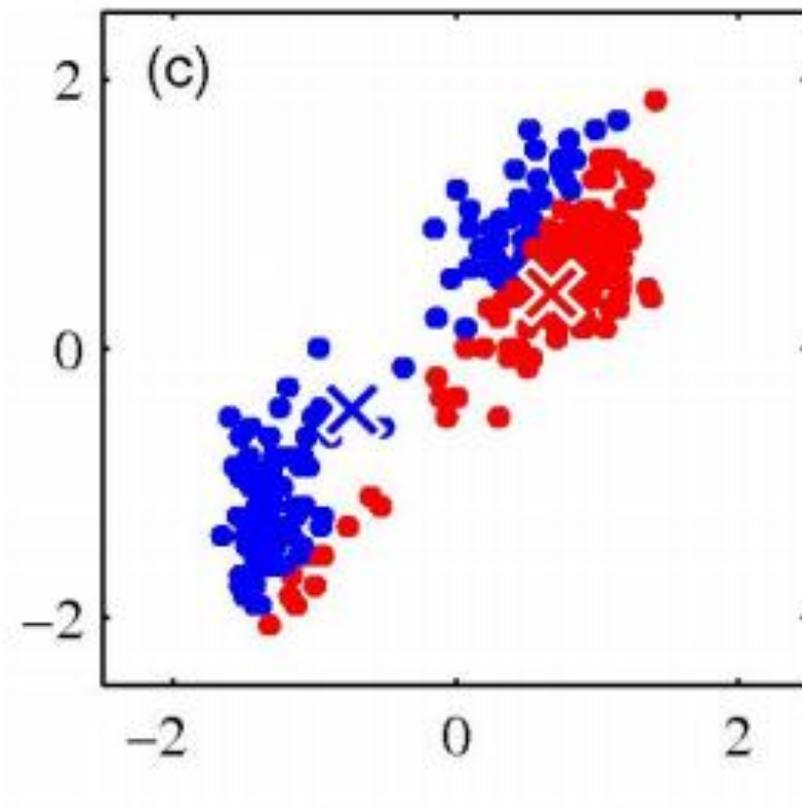
Escolher $k=2$
pontos aleatórios

K-means



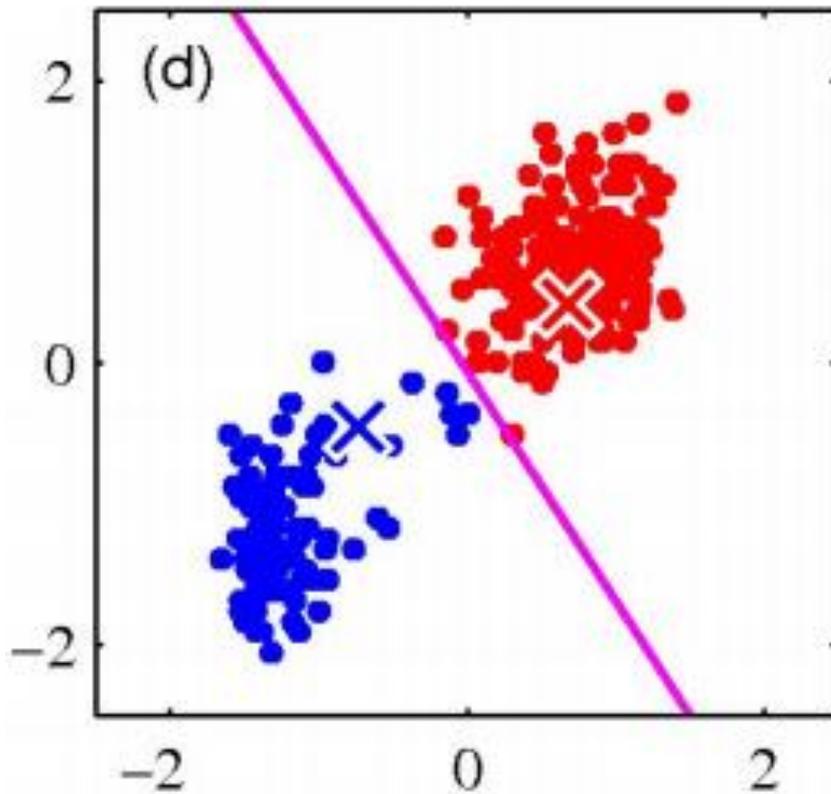
Atribuir os pontos para o centro mais próximo

K-means



Mudar o centro do cluster para a média dos pontos atribuídos

K-means



Repetir até
convergir

Análise Econômico-Cultural

Participatory Sensor Networks as Sensing Layers
SocialCom'14

Setores de NY considerados

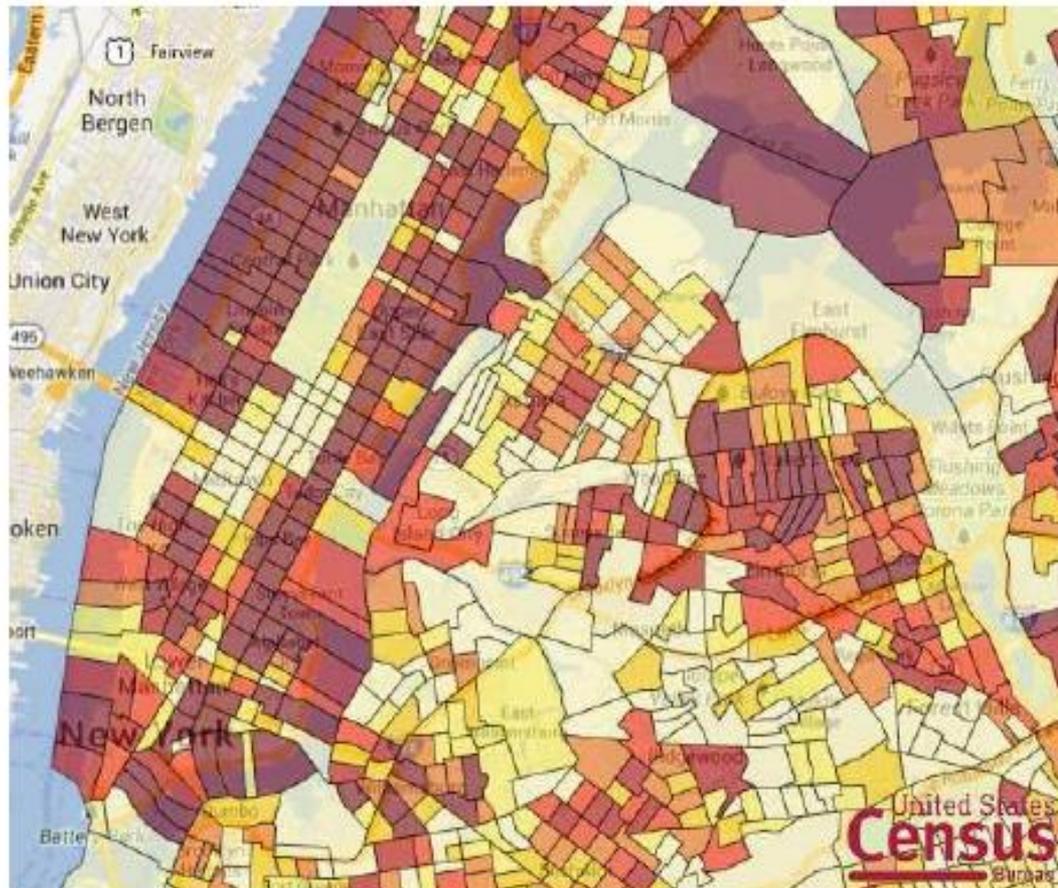
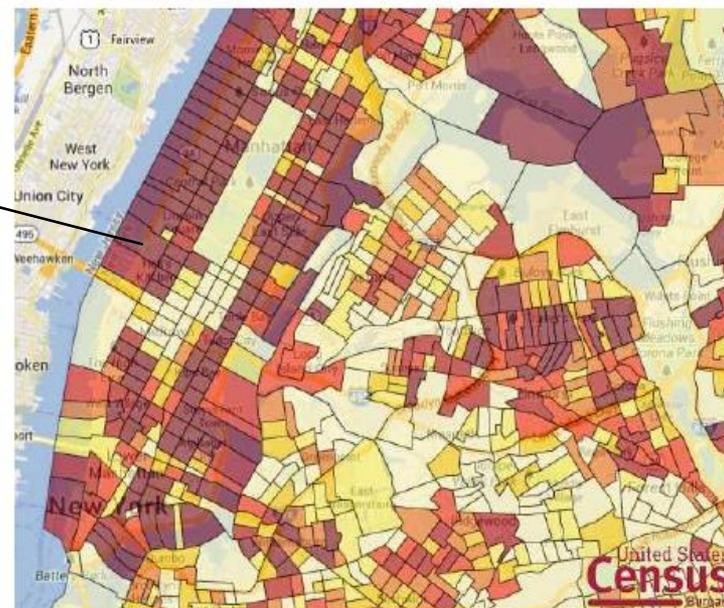


Imagem do censo de NY

Correlação entre sentimento e renda de setores

Dicas de locais e renda média dos habitantes

Sentimento de todas as dicas. Resultando em um valor médio por setor



Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Faixas de sentimento



Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
<25000	0,46 (0,67)	0	73,08	21,15	5,77	0
≥25000 and <50000	0,73 (0,63)	1,23	84,31	12,92	1,23	0,31
≥50000 and <75000	0,81 (0,46)	0,40	93,28	5,93	0,39	0
≥75000 and <100000	0,9 (0,36)	0	96,97	3,03	0	0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
<25000	0,46 (0,67)	0	73,08	21,15	5,77	0
≥25000 and <50000	0,73 (0,63)	1,23	84,31	12,92	1,23	0,31
≥50000 and <75000	0,81 (0,46)	0,40	93,28	5,93	0,39	0
≥75000 and <100000	0,9 (0,36)	0	96,97	3,03	0	0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0



Faixas de renda

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
<25000	0,46 (0,67)	0	73,08	21,15	5,77	0
≥25000 and <50000	0,73 (0,63)	1,23	84,31	12,92	1,23	0,31
≥50000 and <75000	0,81 (0,46)	0,40	93,28	5,93	0,39	0
≥75000 and <100000	0,9 (0,36)	0	96,97	3,03	0	0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0

Setores pobres tendem a ter o pior sentimentos expressados pelos usuários

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
<25000	0,46 (0,67)	0	73,08	21,15	5,77	0
≥25000 and <50000	0,73 (0,63)	1,23	84,31	12,92	1,23	0,31
≥50000 and <75000	0,81 (0,46)	0,40	93,28	5,93	0,39	0
≥75000 and <100000	0,9 (0,36)	0	96,97	3,03	0	0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0

Com o aumento da renda as opiniões tendem a ser melhores

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
<25000	0,46 (0,67)	0	73,08	21,15	5,77	0
≥25000 and <50000	0,73 (0,63)	1,23	84,31	12,92	1,23	0,31
≥50000 and <75000	0,81 (0,46)	0,40	93,28	5,93	0,39	0
≥75000 and <100000	0,9 (0,36)	0	96,97	3,03	0	0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0

↓
Número maior de dicas positivas (faixa de renda alta)

Escala de sentimento: -4 (muito negativa) a +4 (muito positiva)

Group	Mean Sent. (std)	(+3,+4)%	(+1,+2)%	(0)%	(-1,-2)%	(-3,-4)%
						0
≥2500						0,31
≥5000						0
≥7500						0
≥100000	0,87 (0,28)	0,96	98,08	0,96	0	0

Potencial para estudos de desigualdades na qualidade de serviços em cidades

Análise de sentimento: identificar e extrair informações subjetivas de materiais textuais

Classificar a polaridade da opinião expressa pelo autor em positiva, negativa ou neutra

Existem vários métodos

SentiStrength: desenvolvido para textos curtos

Sentimento positivo de força 1 (nenhum sentimento positivo) a 5 (sentimento positivo muito forte)

Sentimento negativo de força -1 (nenhum sentimento negativo) a -5 (sentimento negativo muito forte) para cada texto

Existem vários métodos

SentiStrength: desenvolvido para textos curtos

Sentimento positivo de força 1 (nenhum sentimento positivo) a 5 (sentimento positivo muito forte)

Sentimento negativo de força -1 (nenhum sentimento negativo) a -5 (sentimento negativo muito forte) para cada texto

Positivo + **negativo** = score único

Lista de várias palavras associadas a uma determinada força de sentimento (parte importante)

“bom” possui valor 3

“medo” possui valor -4

	positive, negative
◆ My legs ⁻² <u>ache</u> .	1, -2
◆ You are the ³ <u>coolest</u> .	3, -1
◆ I <u>hate</u> ⁻⁴ Paul but <u>encourage</u> ² him.	2, -4

Alocação de lojas

Geo-Spotting: Mining Online Location-based Services for Optimal Retail Store Placement –
KDD13

Alocação de lojas



Onde instalar uma nova loja para ser mais lucrativa?

Alocação de lojas



Starbucks

1 Penn Plaza (at W 34th St), New York, NY 10001, United States
Coffee Shop



+1 212-736-3001 @starbucks starbucks.com
Report a problem

8.1/10
People like this place

YOUR CHECK-INS	TOTAL PEOPLE	TOTAL CHECK-INS
0	5,940	13,451

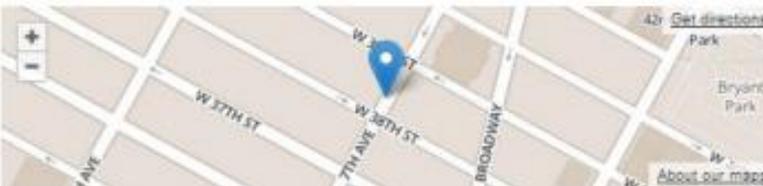
THERE ARE 5 PEOPLE HERE NOW.

The distance between the two places is only few hundred meters



Starbucks

525 Fashion Ave (btw 38th & 39th St), New York, NY 10018, United States
Coffee Shop, Coffee Shop

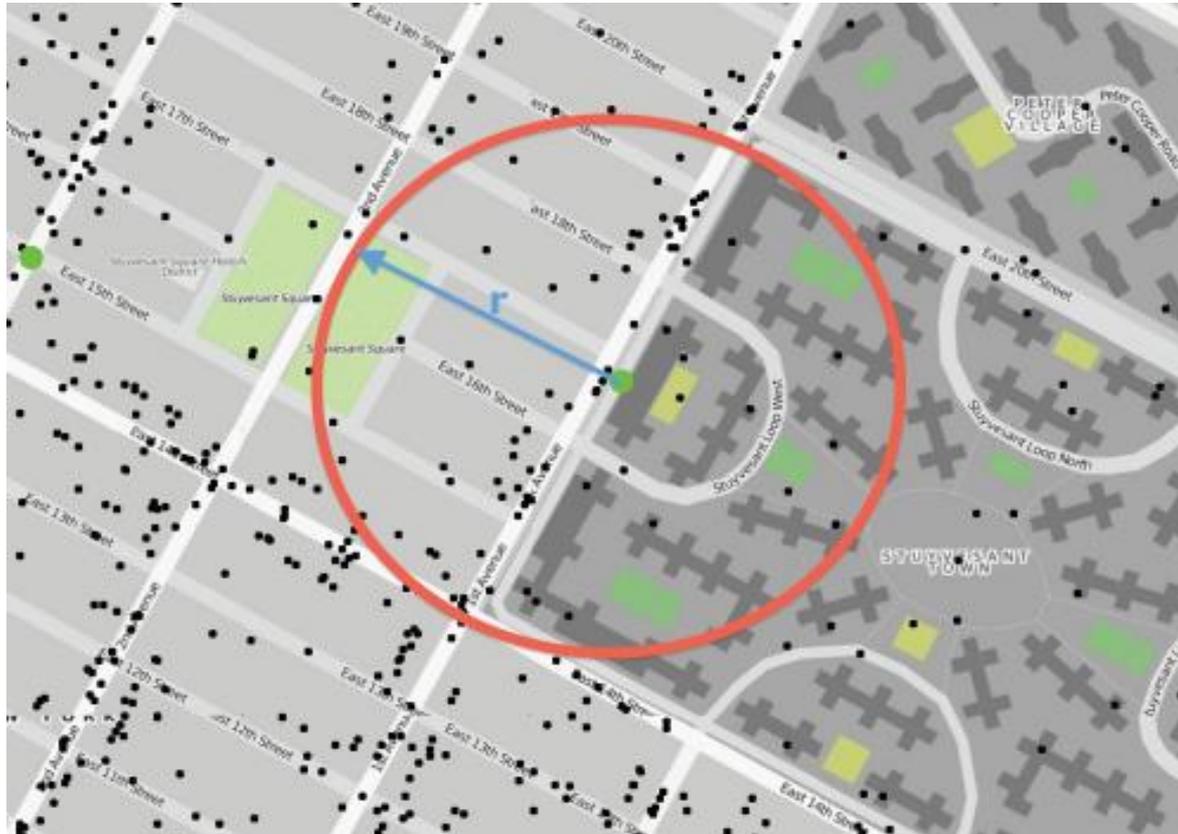


+1 212-869-5273 @starbucks starbucks.com
See the menu
Report a problem

6.3/10
People like this place

YOUR CHECK-INS	TOTAL PEOPLE	TOTAL CHECK-INS
0	2,350	6,557

Mayol. Liquica Z.
48 check-ins in last 60 days



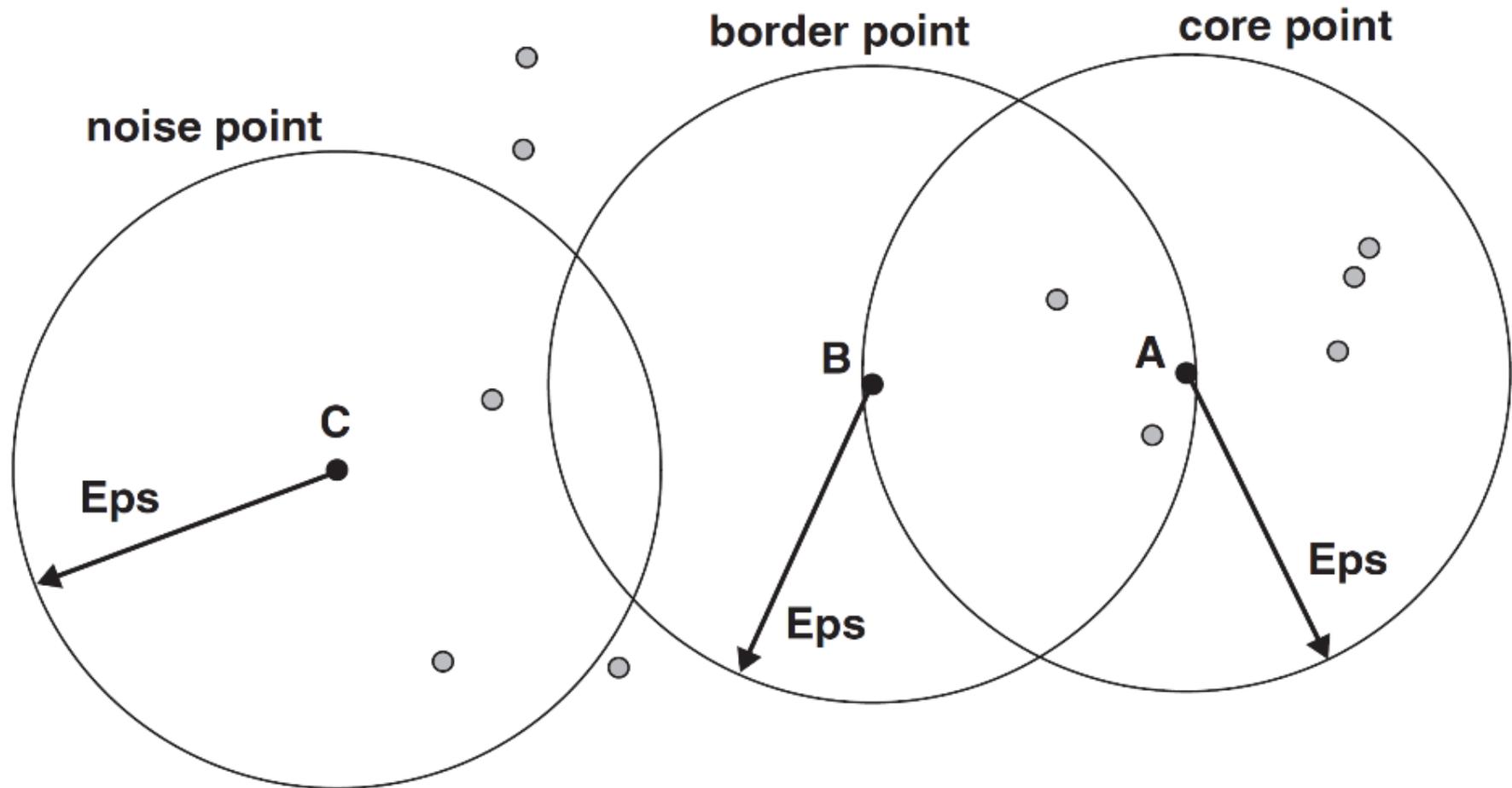
Uma área é descrita por um conjunto de características

DBSCAN é um algoritmo baseado em densidade.

Densidade = número de pontos dentro de um raio especificado (*Eps*)

- **Core point**: atinge um número mínimo de pontos especificados pelo usuário (**minPts**) dentro do raio *Eps*
- **Border point** fica localizado na vizinhança de um *core point* e não é um *core point*
- **Noise point** qualquer outro ponto

MinPts = 7



- Rotular todos os pontos
- Eliminar **noise points**
- Conectar todos os **core points** que estejam dentro da *Eps* um dos outros
- Cada conjunto de **core points** conectados é um grupo
- Atribuir cada **border point** a um grupo de **core points**

Problemas das cidades



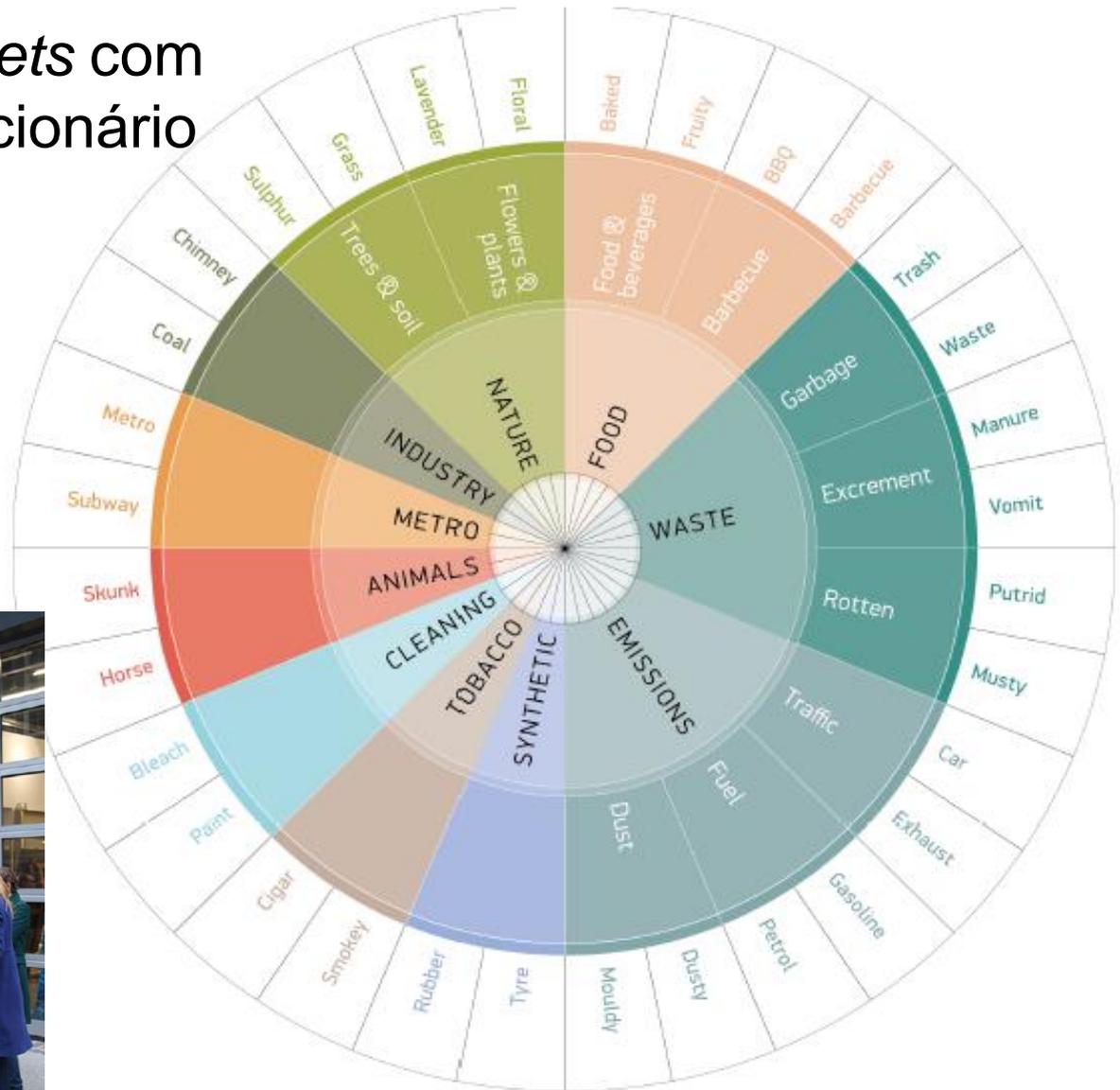
Mapa de Cheiro

Smelly Maps: The Digital Life of Urban Smellscapes
ICWSM'15

Objetivo: mapear os cheiros percebidos em diversas regiões da cidade

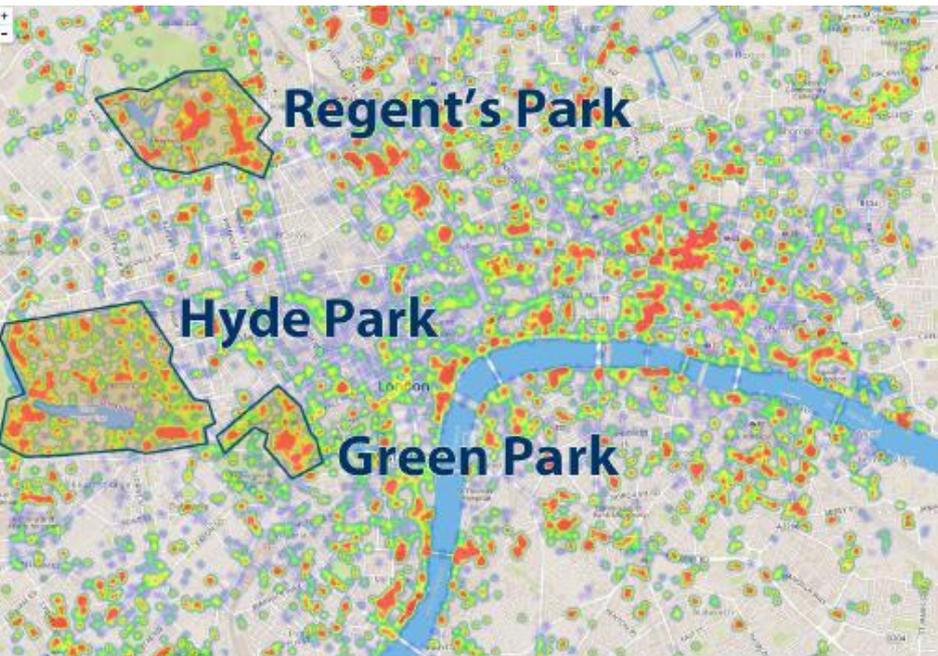
Mapa de cheiro

Combinam *tags* e *tweets* com as palavras de um “dicionário de cheiro” predefinido



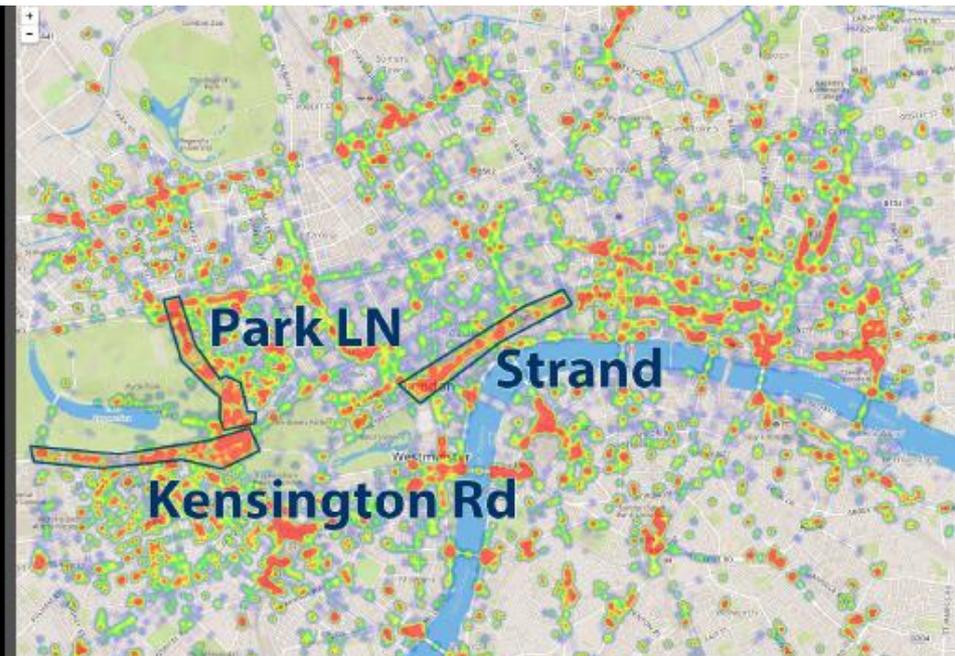
Imagens de Daniele Quercia

Natureza



Áreas verdes da cidade

Emissão gasosa



Ruas com tráfego intenso

Natureza



Emissão gasosa



Pode ser uma nova forma para classificar áreas de acordo com o seu cheiro mais característico

Áreas verdes da cidade

Ruas com tráfego intenso

Mobilidade urbana



Rotas Felizes

The Shortest Path to Happiness: Recommending
Beautiful, Quiet, and Happy Routes in the City
HT '14

Nem sempre o menor caminho é o que gostaríamos de percorrer

Um turista, por exemplo, pode optar por um caminho mais bonito, mesmo que a distância seja um pouco maior

UrbanGems: Crowdsourcing Quiet, Beauty and Happiness

Change Question

Which place do you find more beautiful?

Progress: 5/10



Picture Info



Picture Info

Can't Tell

UrbanGems: Crowdsourcing Quiet, Beauty and Happiness

Change Question

Which place do you find more beautiful?

Progress: 5/10



Picture Info



Picture Info

Can't Tell

Em seguida, os autores traduzem os votos em medidas quantitativas de percepção de localização

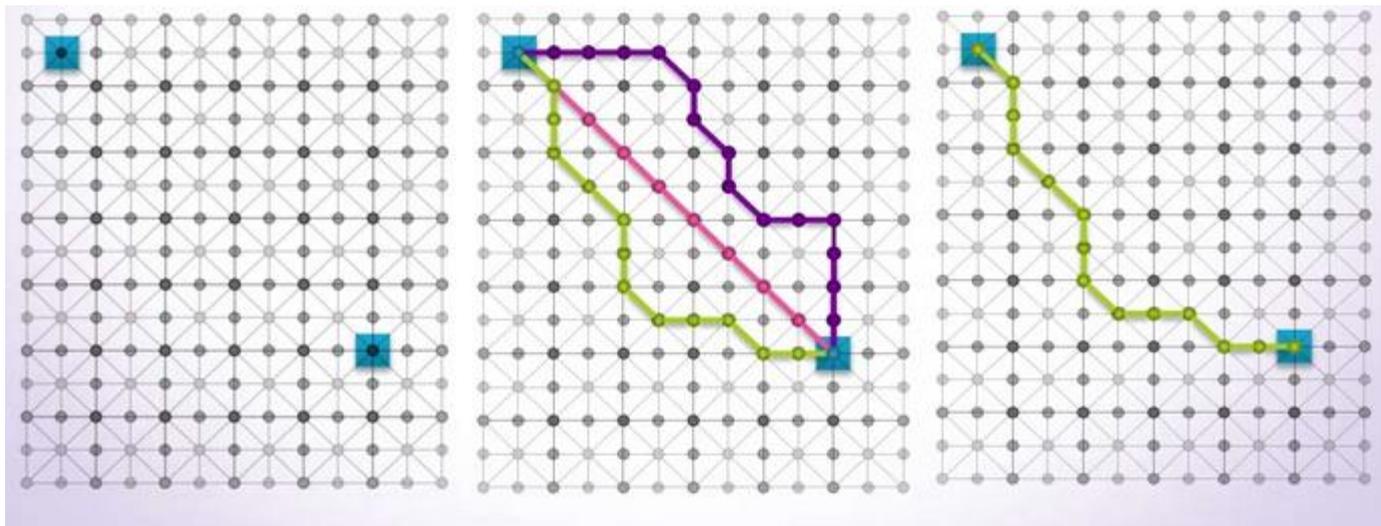


Divide a cidade em um grid
Cada célula tem os *scores* emocionais calculados



Divide a cidade em um grid
Cada célula tem os *scores* emocionais calculados

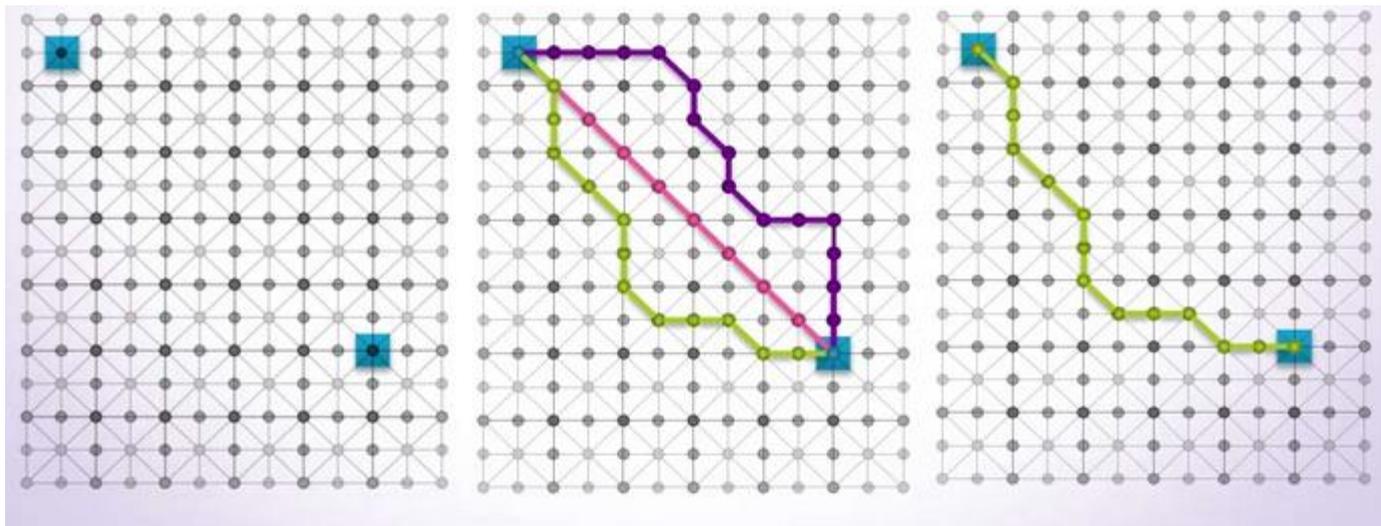
Grafo: nós são as células e arestas conectam vizinhos





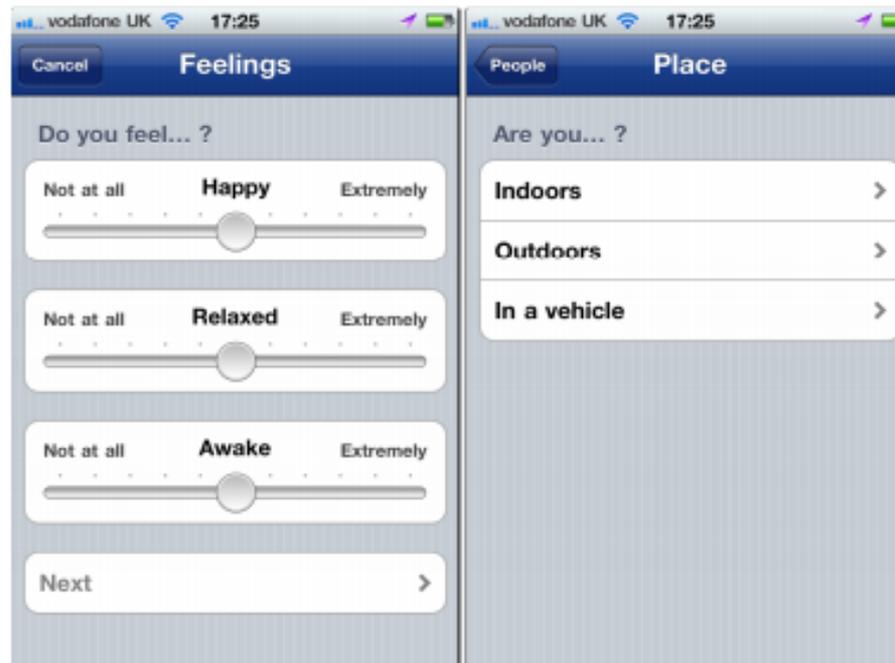
Divide a cidade em um grid
Cada célula tem os *scores* emocionais calculados

Grafo: nós são as células e arestas conectam vizinhos
Calculam as k-rotas mais curtas



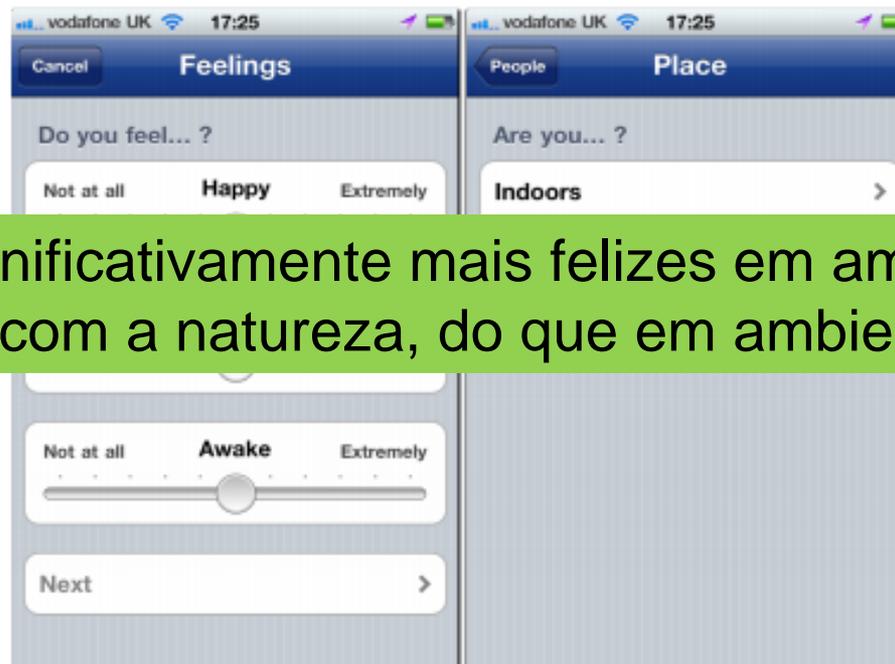


Usuários informaram esses dados em diferentes situações



Happiness is greater in natural environments - 2012

Usuários informaram esses dados em diferentes situações



Usuários são significativamente mais felizes em ambientes externos, em contato com a natureza, do que em ambientes urbanos

Happiness is greater in natural environments - 2012

Cidades e centralidades

Revealing the City that We Cannot See
ACM TOIT'14



BH – (dia)

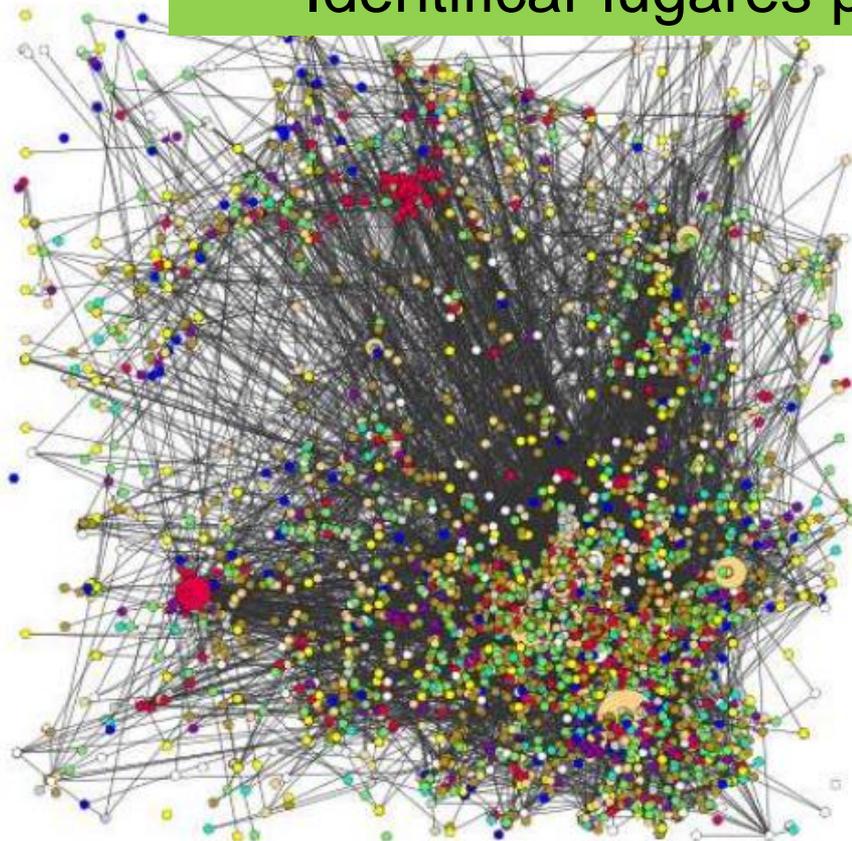
NYC – (dia)

Centralidade de grau

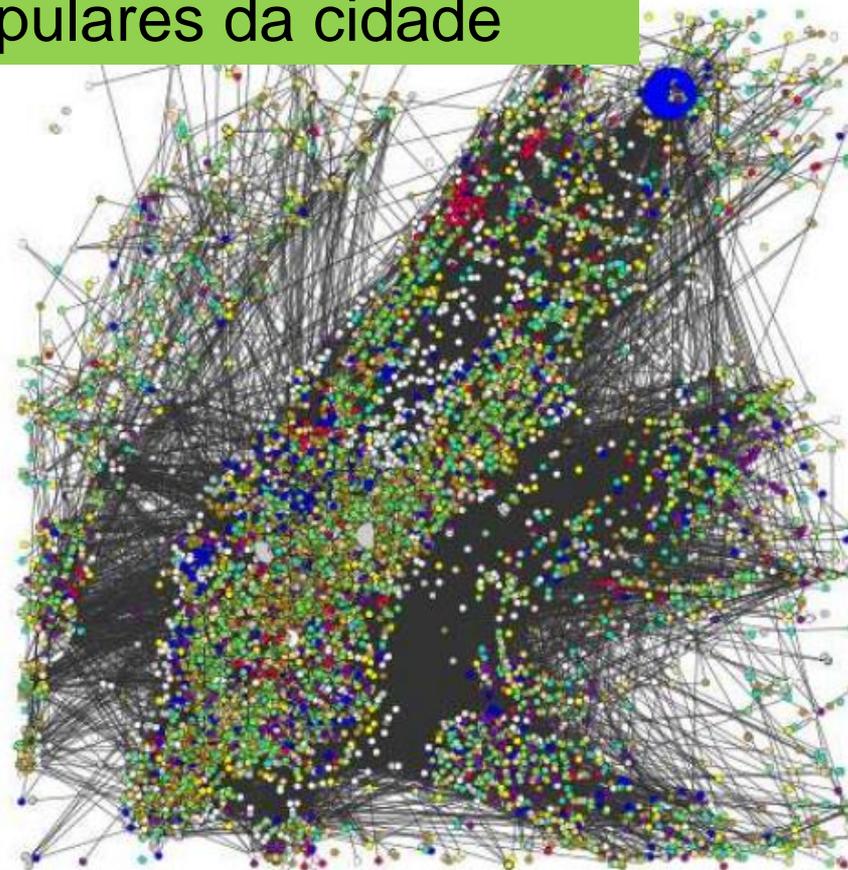
$$C_D(v) = \text{deg}(v)$$

A centralidade do nó v é o seu número de grau

Identificar lugares populares da cidade

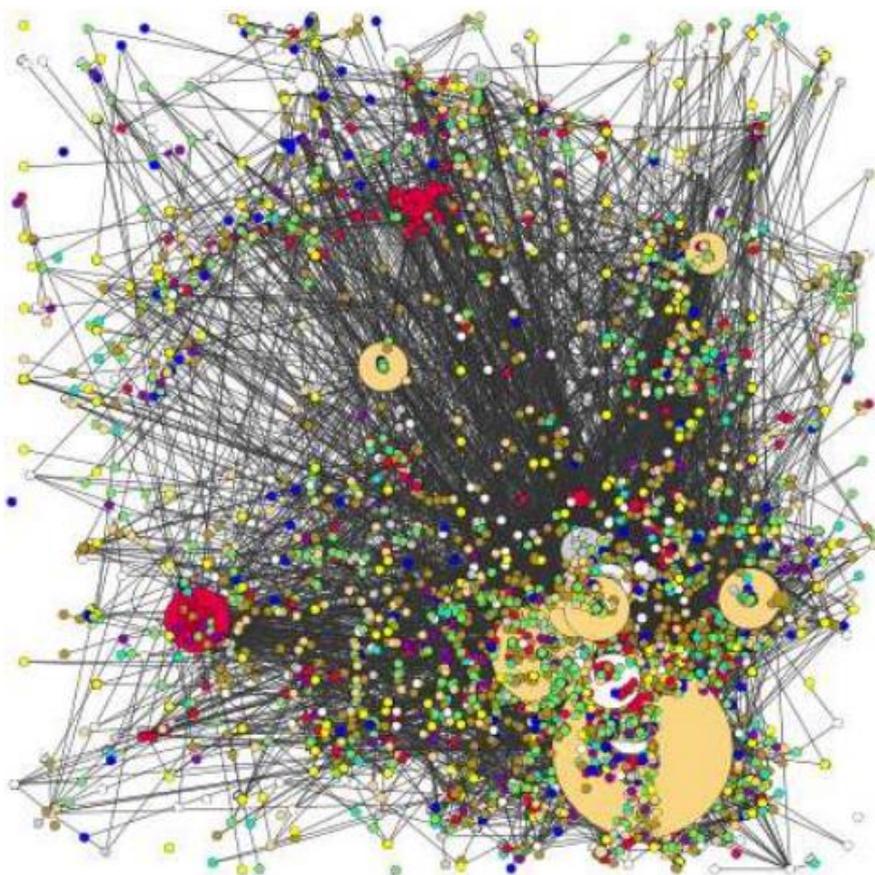


BH – (dia)

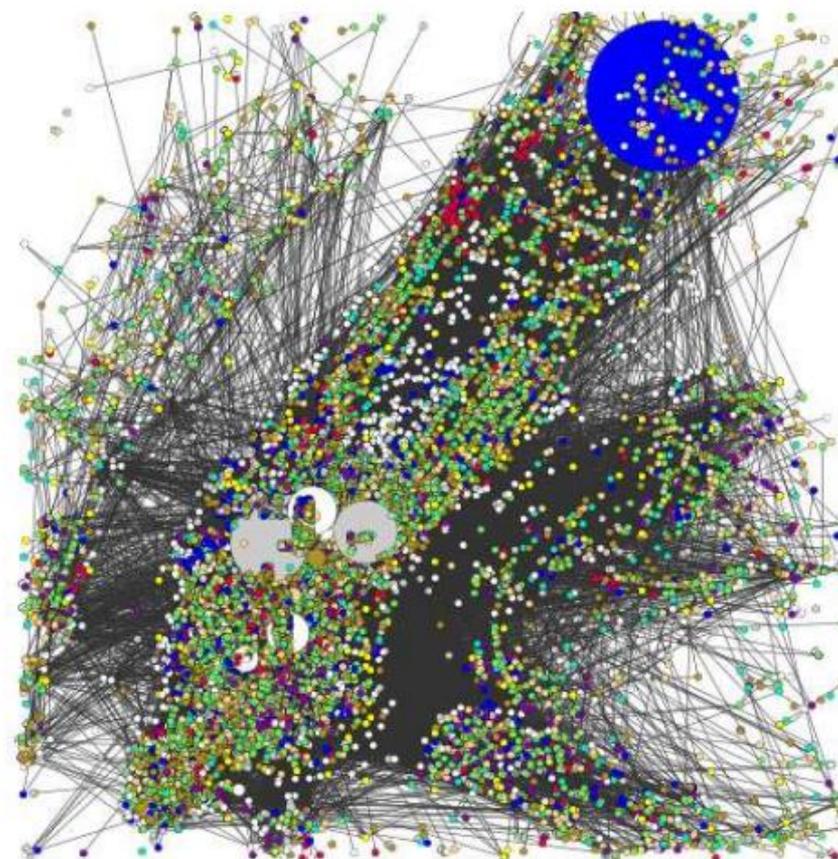


NYC – (dia)

Centralidade de grau



BH – (dia)



NYC – (dia)

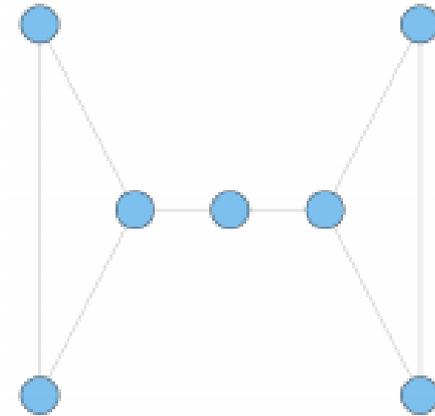
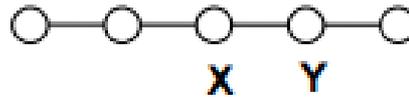
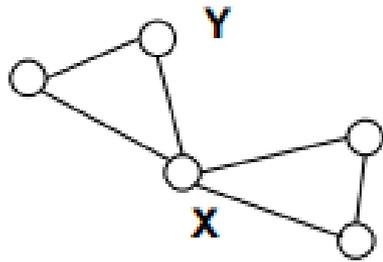
Centralidade de intermediação

$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

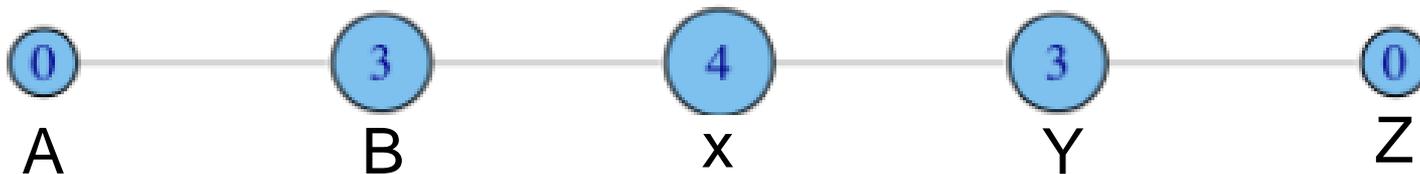
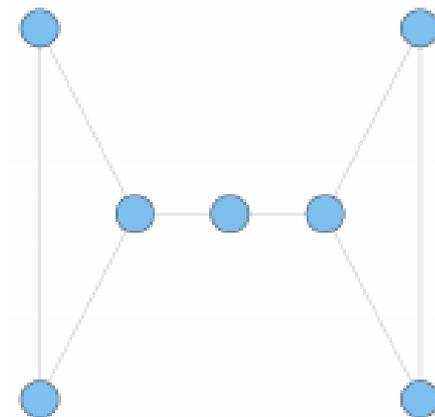
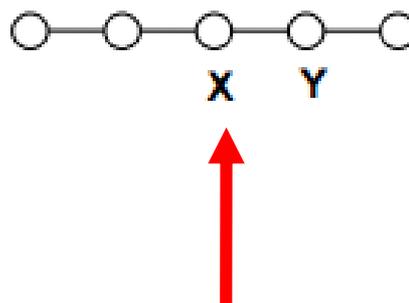
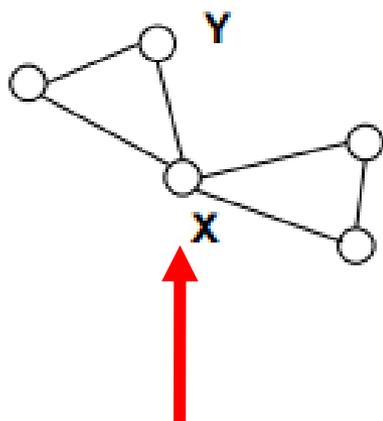
σ_{st} é o número total de menores caminhos do nó s para o nó t

$\sigma_{st}(v)$ é o número de menores caminhos que passam por v

Quem tem o maior valor de intermediação x ou y?



Quem tem o maior valor de intermediação x ou y?



A não fica entre nenhum outro par

B fica entre (A-X), (A-Y), (A-Z)

X fica entre (A-Y), (A-Z), (B-Y), (B-Z)

TOP 10 locais de acordo com a centralidade de intermediação

4b0fcfabf964a520246523e3	Pátio Savassi	0.0967193168
4b4604f9f964a520521426e3	Diamond Mall	0.0544934973
4bace766f964a520c1173be3	Shopping Cidade	0.0350708903
4bc5e72651b376b093981a6f	PUC Minas	0.0347481753
4ccc7075ee23a14301171fa8	Boulevard Shopping	0.0330330587
4b6973f2f964a520eda22be3	Praça da Liberdade	0.0297484327
4b5ec87ff964a520909829e3	Mercado Central	0.0275445564
4babc507f964a52005c83ae3	Shopping Del Rey	0.0267350643
4b0fcf66f964a520206523e3	Savassi	0.0259948311
4b5a3ad1f964a520fdb528e3	Terminal Rodoviário de Belo Horizonte	0.0242963552

TOP 10 locais de acordo com a centralidade de intermediação

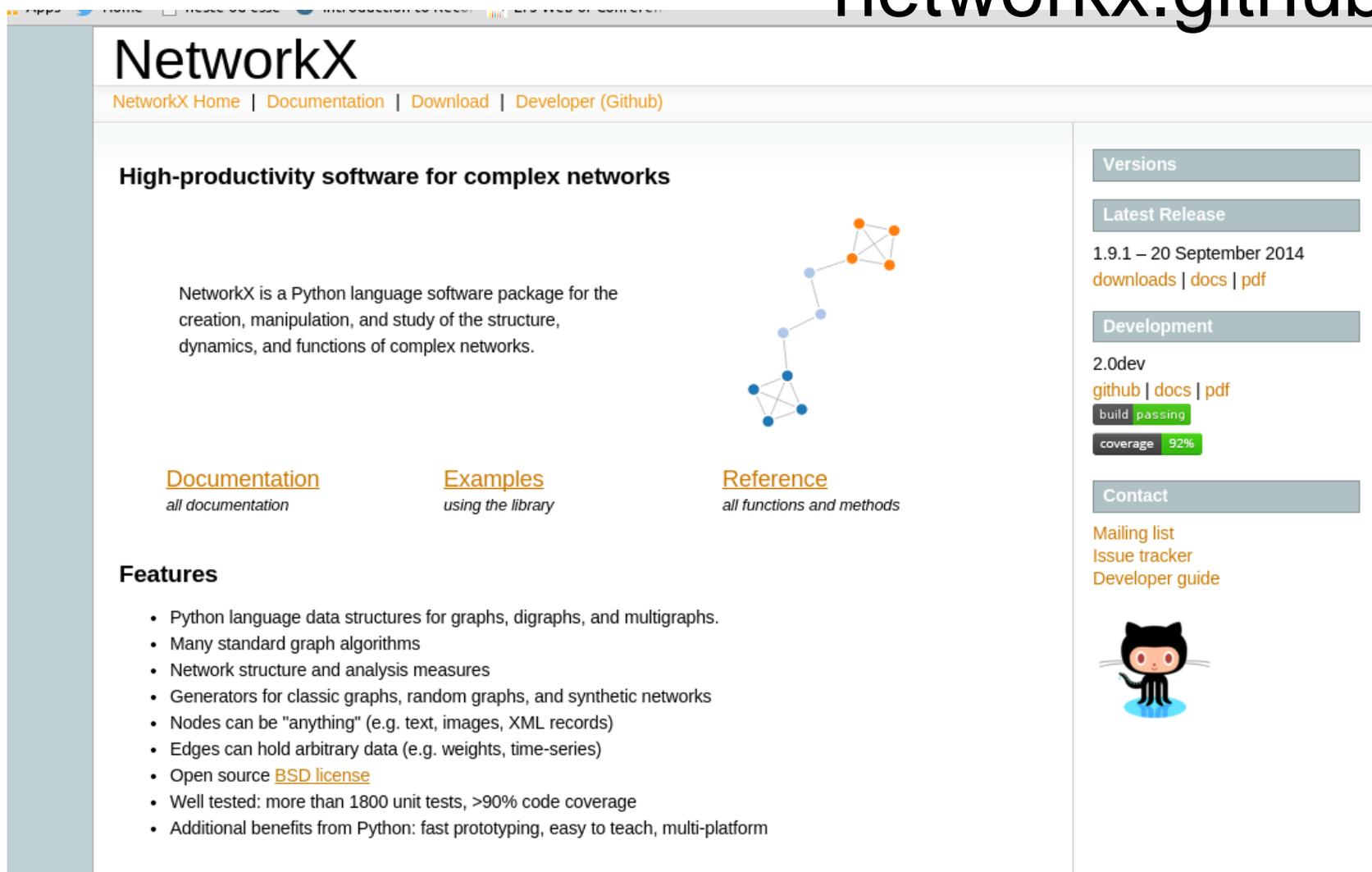
4b0fcfabf964a520246523e3	Pátio Savassi	0.0967193168
4b4604f9f964a520521426e3	Diamond Mall	0.0544934973
4bace766f964a520c1173be3	Shopping Cidade	0.0350708903

Quanto maior a intermediação, maior a probabilidade de que um usuário passe por este local

4b6973f2f964a520eda22be3	Praça da Liberdade	0.0297484327
4b5ec87ff964a520909829e3	Mercado Central	0.0275445564
4babc507f964a52005c83ae3	Shopping Del Rey	0.0267350643
4b0fcf66f964a520206523e3	Savassi	0.0259948311
4b5a3ad1f964a520fdb528e3	Terminal Rodoviário de Belo Horizonte	0.0242963552

TOP 10 locais de acordo com a centralidade de intermediação

4b0fcfabf964a520246523e3	Pátio Savassi	0.0967193168
4b4604f9f964a520521426e3	Diamond Mall	0.0544934973
4bace76	Estratégico para fazer parcerias entre empresas	0.04708903
4bc5e72	Por exemplo, propaganda para direcionar consumidores para outros negócios independentes	0.04481753
4ccc707		0.0330587
4b6973f		0.04484327
4b5ec87ff964a520909829e3	Mercado Central	0.0275445564
4babc507f964a52005c83ae3	Shopping Del Rey	0.0267350643
4b0fcf66f964a520206523e3	Savassi	0.0259948311
4b5a3ad1f964a520fdb528e3	Terminal Rodoviário de Belo Horizonte	0.0242963552



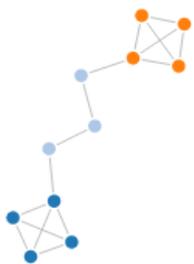
The screenshot shows the NetworkX website homepage. At the top, there's a navigation bar with links for 'NetworkX Home', 'Documentation', 'Download', and 'Developer (Github)'. The main heading is 'NetworkX', followed by the tagline 'High-productivity software for complex networks'. A central text block describes NetworkX as a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks. To the right of this text is a network graph visualization with blue and orange nodes. Below the text are three columns of links: 'Documentation' (all documentation), 'Examples' (using the library), and 'Reference' (all functions and methods). A 'Features' section lists various capabilities like Python language data structures, standard graph algorithms, network structure analysis, generators for different graph types, and data handling on edges. On the right side, there are sections for 'Versions' (Latest Release: 1.9.1 - 20 September 2014), 'Development' (2.0dev, build passing, coverage 92%), and 'Contact' (Mailing list, Issue tracker, Developer guide). A GitHub logo is also present at the bottom right.

NetworkX

[NetworkX Home](#) | [Documentation](#) | [Download](#) | [Developer \(Github\)](#)

High-productivity software for complex networks

NetworkX is a Python language software package for the creation, manipulation, and study of the structure, dynamics, and functions of complex networks.



[Documentation](#)
all documentation

[Examples](#)
using the library

[Reference](#)
all functions and methods

Features

- Python language data structures for graphs, digraphs, and multigraphs.
- Many standard graph algorithms
- Network structure and analysis measures
- Generators for classic graphs, random graphs, and synthetic networks
- Nodes can be "anything" (e.g. text, images, XML records)
- Edges can hold arbitrary data (e.g. weights, time-series)
- Open source [BSD license](#)
- Well tested: more than 1800 unit tests, >90% code coverage
- Additional benefits from Python: fast prototyping, easy to teach, multi-platform

Versions

Latest Release

1.9.1 – 20 September 2014
[downloads](#) | [docs](#) | [pdf](#)

Development

2.0dev
[github](#) | [docs](#) | [pdf](#)

build passing

coverage 92%

Contact

[Mailing list](#)
[Issue tracker](#)
[Developer guide](#)

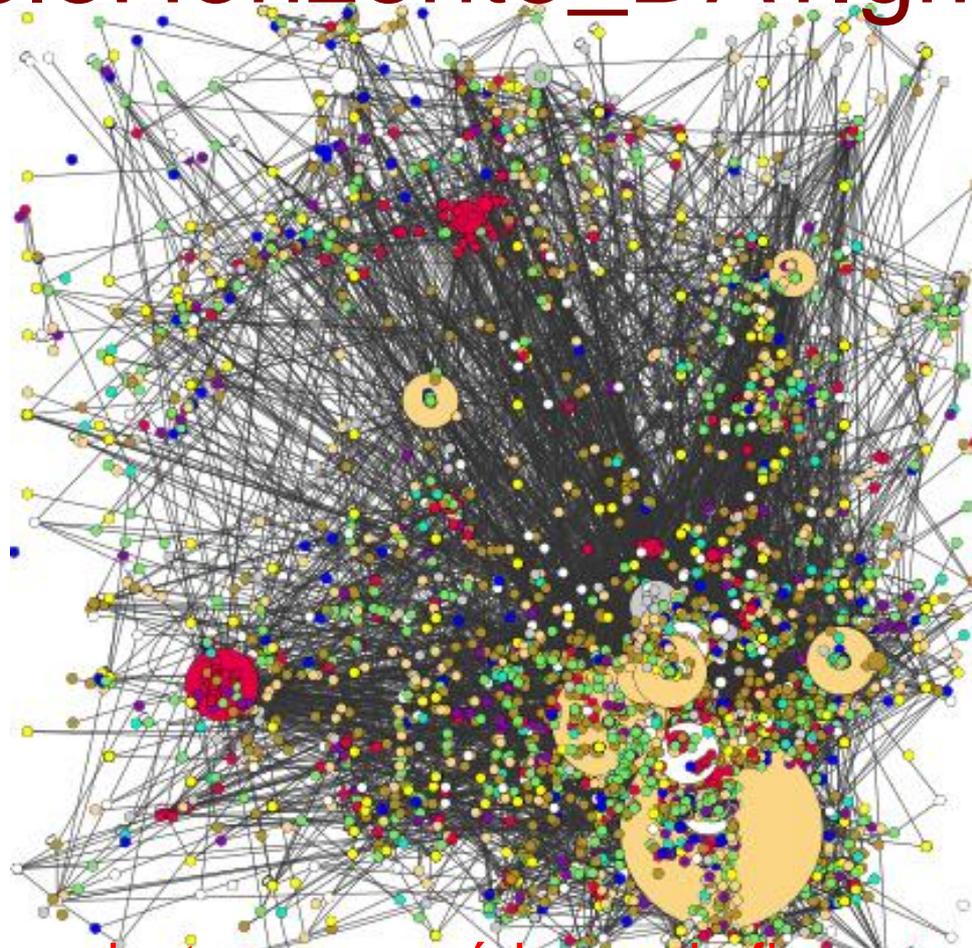


```
1 import networkx as nx
2
3 G=nx.Graph()
4
5 G.add_node("A")
6 G.add_nodes_from(["B","C"])
7
8 G.add_edge("A","B")
9
10 print G.number_of_nodes()
11 print G.number_of_edges()
12
13 #G.add_node(1, time='5pm')
14
15 print G.nodes(data=True)
16
17
18 G.add_nodes_from(["D","E","F","G","H","I","J","K","L"])
19 G.add_edge("C","D")
20 G.add_edge("D","E")
21 G.add_edge("A","C")
22 G.add_edge("I","E")
23 G.add_edge("J","D")
24 G.add_edge("H","D")
25
26
27 nx.write_gml(G,"test.gml")
```

Códigos de exemplo
disponíveis

O grafo de transição urbana: BeloHorizonte_DAY.gml

Cytoscape



Com os ajustes necessários pode ficar nesse estilo