

Escola Nacional de
Ciências Estatísticas



Instituto Brasileiro de Geografia e Estatística



Modelos de redes sociais e técnicas de visualização para dados de Migração

Gustavo da Silva Ferreira

10/Novembro/2016

Redes Sociais

- ▶ Uma rede social se estrutura por um conjunto de atores ou **nós** (pessoas, objetos, locais ou eventos) ligados por uma **relação**;
- ▶ Na teoria dos grafos, estes elementos são conhecidos como **vértices** e **arestas**;
- ▶ Os objetivos da análise de redes sociais abrangem a identificação de nós com papel **importante**, a sumarização, a identificação do **padrão geral** e da existência de **grupos subjacentes** na rede;

Redes Sociais e o Movimento de Pessoas

- ▶ A identificação e a análise dos fluxos migratórios têm sido fruto de inúmeros estudos;
- ▶ O mesmo acontece na análise de movimentos pendulares (moradia-trabalho) nas grandes metrópoles;
- ▶ Causas costumam ser explicadas por teorias sociais e econômicas, aproveitando-se a grande quantidade de dados disponíveis;

Visualização de fluxos

- ▶ A visualização de fluxos migratórios usualmente é realizada por meio de tabelas com estatísticas de saldos migratórios e com mapas plotando as diferentes direções dos fluxos existentes;
- ▶ Alguns autores têm se dedicado a criar formas alternativas para uma melhor visualização e análise dos dados (Aigner et al., 2007, Guo, 2009, Sander et al., 2014, entre outros).

Pergunta

É possível utilizar modelos de espaços latentes para redes sociais a fim de modelar dados de migração e de movimentos de pessoas?

Resposta: SIM!

Modelos para redes sociais

- ▶ Os dados costumam ser organizados em uma matriz **Y** simétrica, de dimensão $n \times n$, onde cada elemento indica a existência (ou não) da relação social entre dois indivíduos da rede;
- ▶ Alternativamente, esta matriz pode conter o grau da relação social, mensurado pelo **número de conexões**, existente entre dois indivíduos da rede;
- ▶ Existem ainda modelos mais complexos que incorporam também a direção da relação, permitindo que **Y não seja uma matriz simétrica.**

Modelos para redes sociais

- ▶ Em geral Y_{ij} é modelado por uma distribuição **Bernoulli** ou **Poisson**, onde o interesse reside em $E[Y_{ij}]$;
- ▶ Esta média é usualmente modelada em função de efeitos aleatórios associados a cada um dos indivíduos da rede;
- ▶ O processo de inferência geralmente é baseado nos métodos de MV e MCMC.

1. Modelo de intensidade de conexões

$$Y_{ij} \sim \text{Bernoulli}(\mu_{ij}), i < j$$

$$\log\left(\frac{\mu_{ij}}{1 - \mu_{ij}}\right) = a_i + a_j$$

onde Y_{ij} é a variável indicadora de presença ou ausência de relação entre os elementos i e j da rede social.

Formulação alternativa

ou

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}), i < j$$

$$\log(\mu_{ij}) = a_i + a_j$$

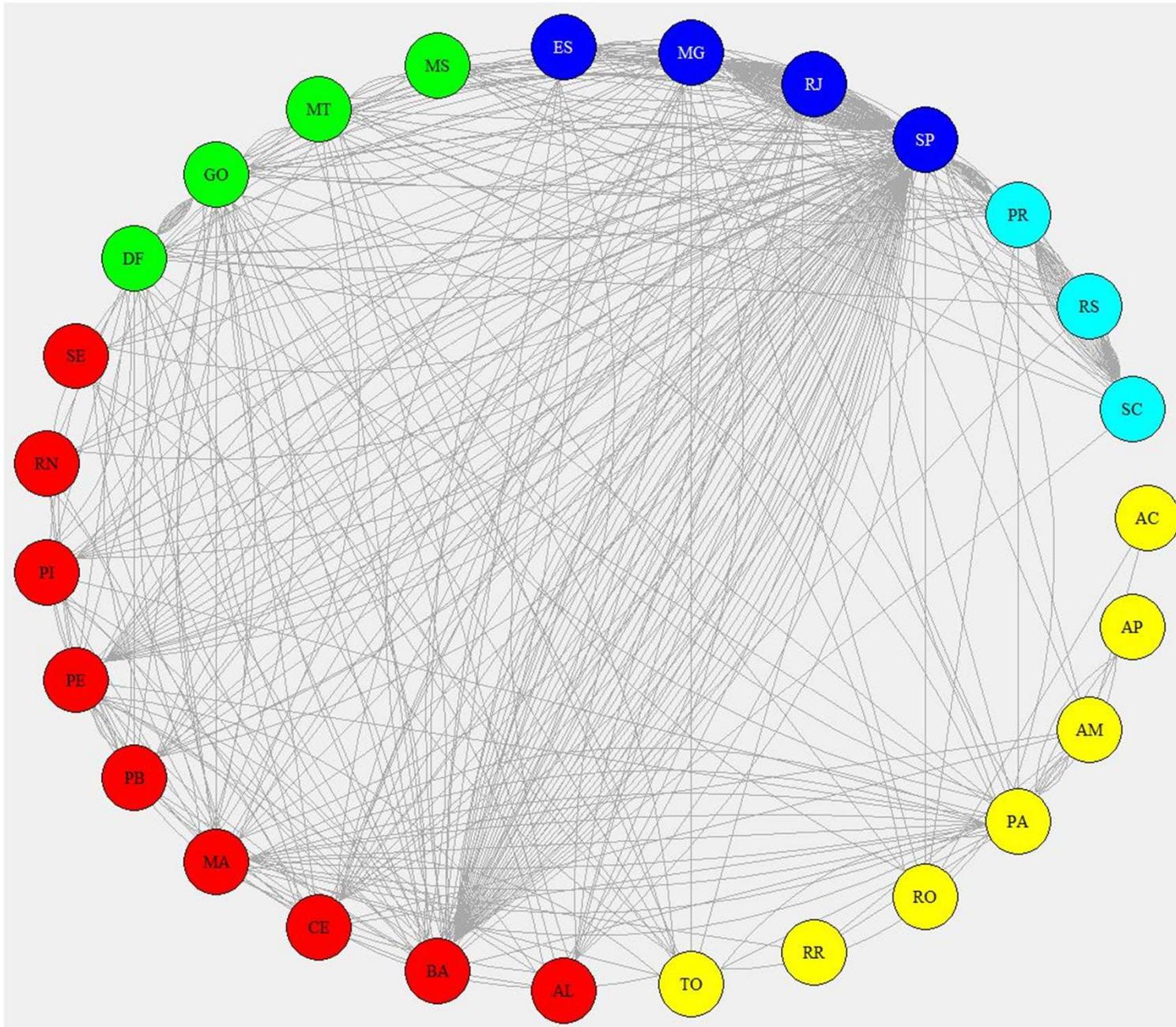
onde Y_{ij} é a variável que contabiliza o número de conexões entre os elementos i e j da rede social.

Observações

- ▶ Fácil implementação e bom ajuste;
- ▶ O efeito a_i é interpretado como um índice de **propensão para conexões (caso Bernoulli)** ou como um índice de **intensidade de conexões (caso Poisson)** na rede social.

Aplicação I: Fluxos migratórios entre UF's no Brasil

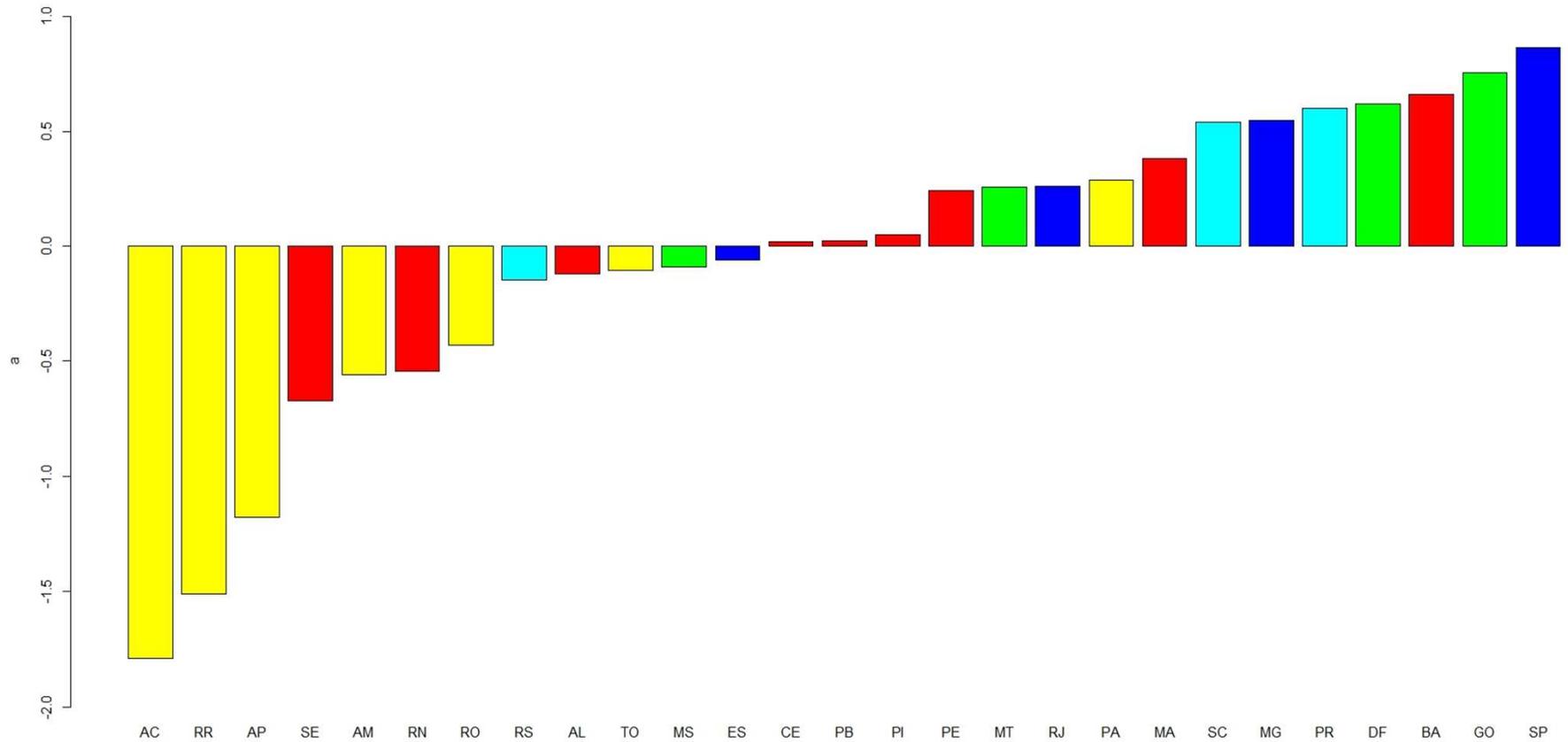
- ▶ Dados: fluxo migratório (número de pessoas) entre Unidades Federativas (UF) no Brasil ocorridos entre os anos de 2005 e 2010;
- ▶ Fonte: Censo 2010 – IBGE;
- ▶ Dados brutos podem ser representados na forma de um grafo não-orientado.



Fluxos migratórios entre UF's no Brasil

- ▶ Cada aresta no grafo representa **10 mil pessoas** migrando (em qualquer direção);
- ▶ Muito **difícil** extrair informações deste grafo;
- ▶ Este grafo representa apenas **uma** dentre **infinitas** formas de representação destes dados.
- ▶ Um modelo de intensidade de conexões (fluxos) foi ajustado aos dados.

Estimativas de a_1, \dots, a_{27} via MV



Fluxos migratórios entre UF's no Brasil

- ▶ Os efeitos de intensidade de fluxo evidenciam o maior deslocamento de pessoas envolvendo os Estados de SP, GO e BA;
- ▶ Por outro lado, os Estados da Região Norte apresentam menores efeitos de intensidade de fluxo;
- ▶ Note que este modelo não considera o efeito das **populações** envolvidas e também não considera a **direção** dos fluxos!

2. Modelo de Distâncias Latentes (Hoff et al, 2002 e Handcock et al, 2007)

Considera que cada um dos n elementos da rede assume uma posição em um espaço latente definido em \mathfrak{R}^d , i.e.

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}), i < j$$

$$\log(\mu_{ij}) = \theta - |a_i - a_j|$$

onde $|a_i - a_j|$ representa a distância euclidiana entre as posições a_i e a_j no espaço latente.

Modelo de Distâncias Latentes para Fluxos migratórios entre UF's no Brasil

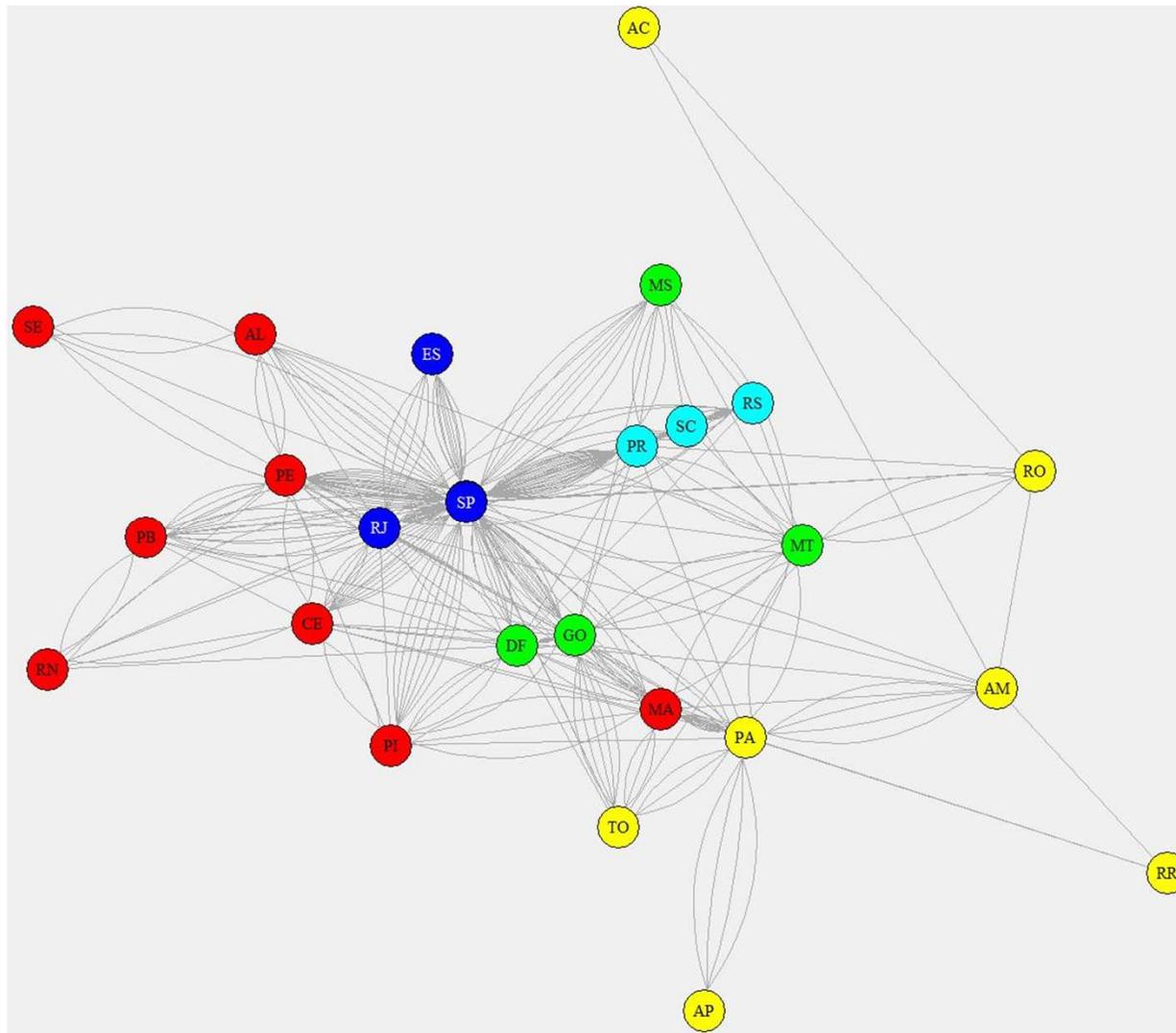
$$Y_{ij} \sim \text{Poisson}(\mu_{ij}), i < j$$

$$\mu_{ij} = e_{ij}\lambda_{ij}$$

$$\log(\lambda_{ij}) = \theta - |a_i - a_j|$$

onde e_{ij} é o fluxo migratório esperado entre as UF's i e j
caso todos os Estados tivessem o mesmo padrão de fluxo.

Posições das UF's no espaço latente de migrações no período 2005-2010 estimada via MV.



3. Escalonamento Multidimensional

- ▶ O Escalonamento Multidimensional (EMS) é uma técnica que utiliza a proximidade entre objetos para criar uma representação espacial dos mesmos a partir de uma matriz de dissimilaridades \mathbf{D} .
- ▶ O objetivo é encontrar uma configuração de pontos em um espaço latente de tal forma que as suas coordenadas produzam uma matriz de distâncias Euclidianas tão próxima quanto possível de \mathbf{D} .

Aplicação II: Pendularidade na Região Metropolinada do Rio de Janeiro (Mattos, 2016)

- ▶ **Dados:** O movimento captado através das perguntas de deslocamento para o trabalho (Município em que trabalha e se retorna do trabalho para a casa diariamente).
- ▶ **Fonte:** Censo 2010 – IBGE;
- ▶ Atualmente a Região Metropolitana do Rio de Janeiro (RMRJ) é composta por 21 municípios, sendo a segunda maior área metropolitana do Brasil.

Pendularidade na Região Metropolinada do Rio de Janeiro

- ▶ Neste caso temos interesse em analisar os fluxos pendulares considerando a direção dos mesmos;
- ▶ Assim, a matriz com os fluxos pendulares Y não será simétrica!
- ▶ O objetivo é encontrar padrões de **chegada** e de **saída** de trabalhadores em cada município.

Ilustração do método (Mattos, 2016)

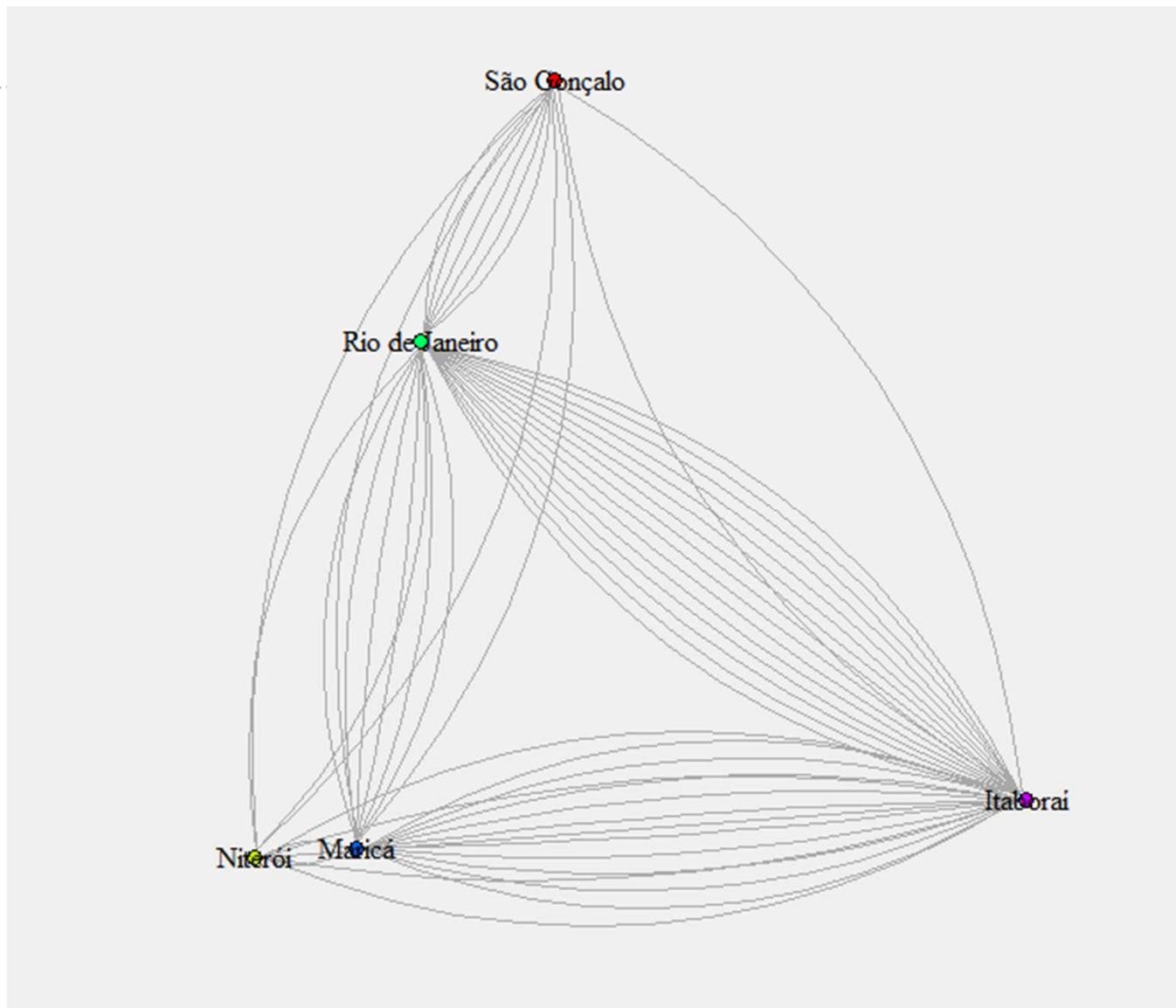
Deslocamento diário de trabalhadores entre 5 municípios

| | | Município de trabalho | | | | |
|-------------------------|----------------|-----------------------|---------|----------------|--------|----------|
| | | São Gonçalo | Niterói | Rio de Janeiro | Maricá | Itaboraí |
| Município de Residência | São Gonçalo | 0 | 20 | 60 | 3 | 16 |
| | Niterói | 10 | 0 | 20 | 8 | 50 |
| | Rio de Janeiro | 2 | 3 | 0 | 60 | 80 |
| | Maricá | 20 | 20 | 50 | 0 | 100 |
| | Itaboraí | 2 | 50 | 150 | 70 | 0 |

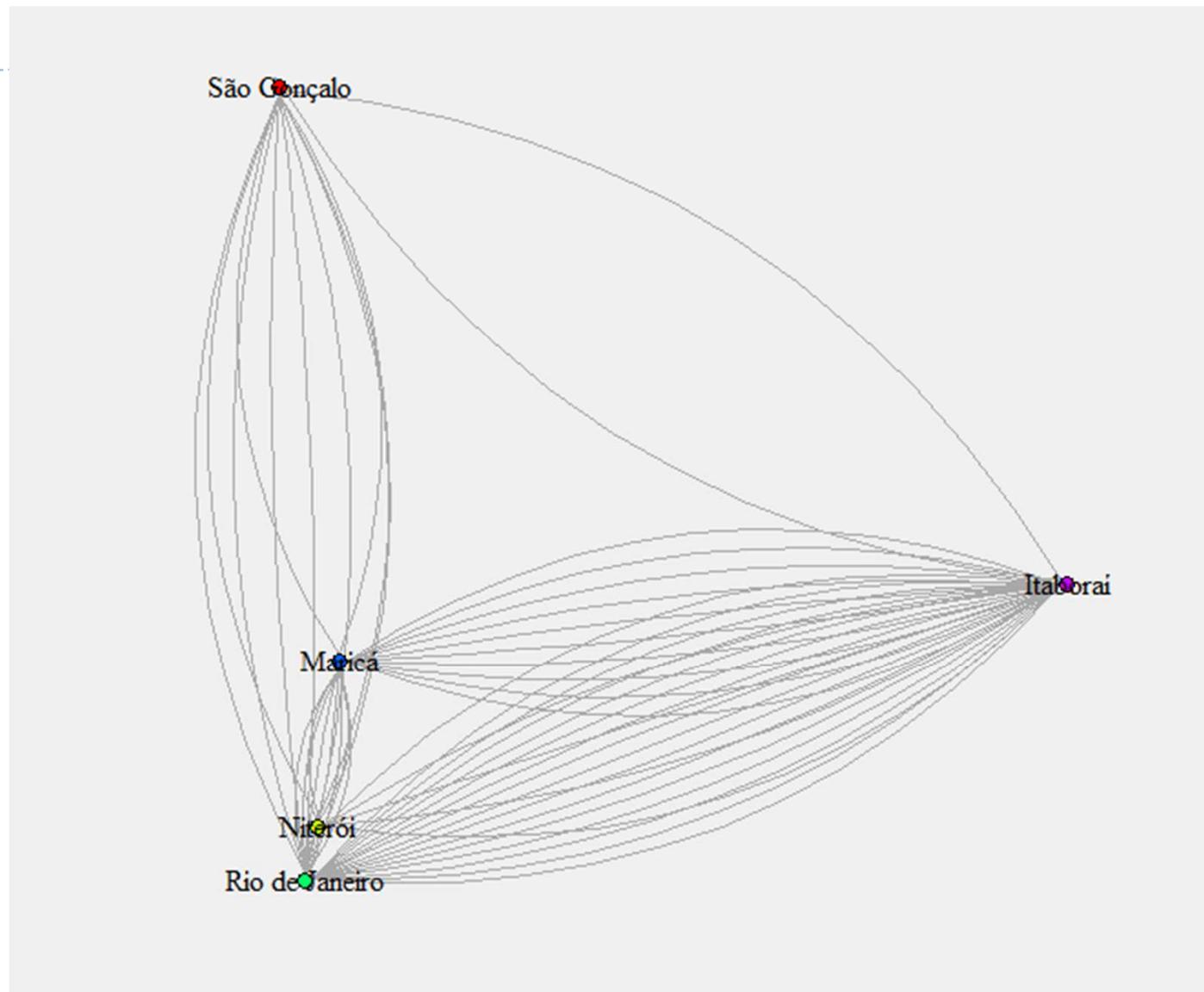
Tamanho populacional dos 5 municípios

| Municípios | População |
|----------------|-----------|
| São Gonçalo | 2000 |
| Niterói | 1000 |
| Rio de Janeiro | 3000 |
| Maricá | 2000 |
| Itaboraí | 2100 |

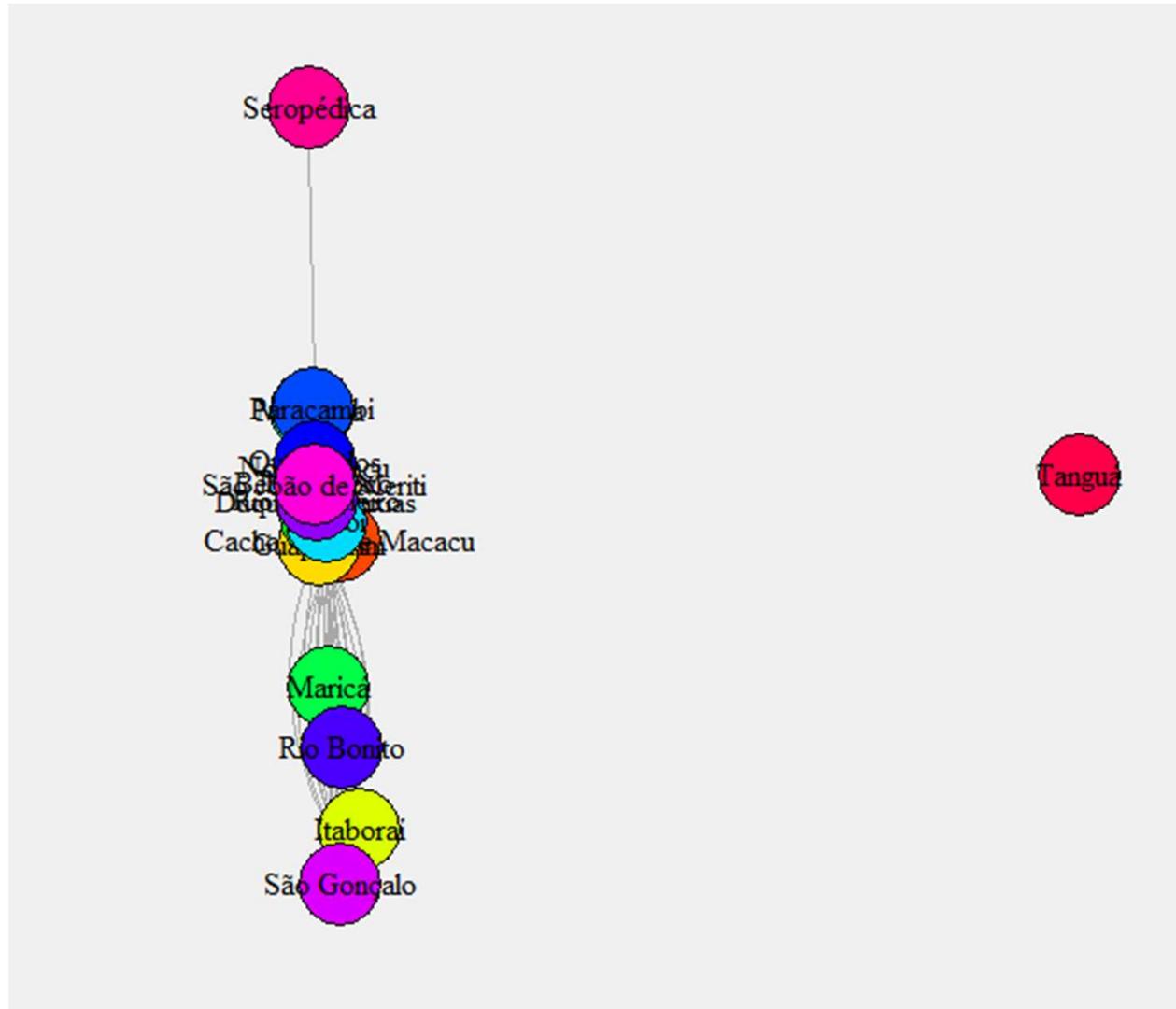
Representação espacial da movimentação de **saída**



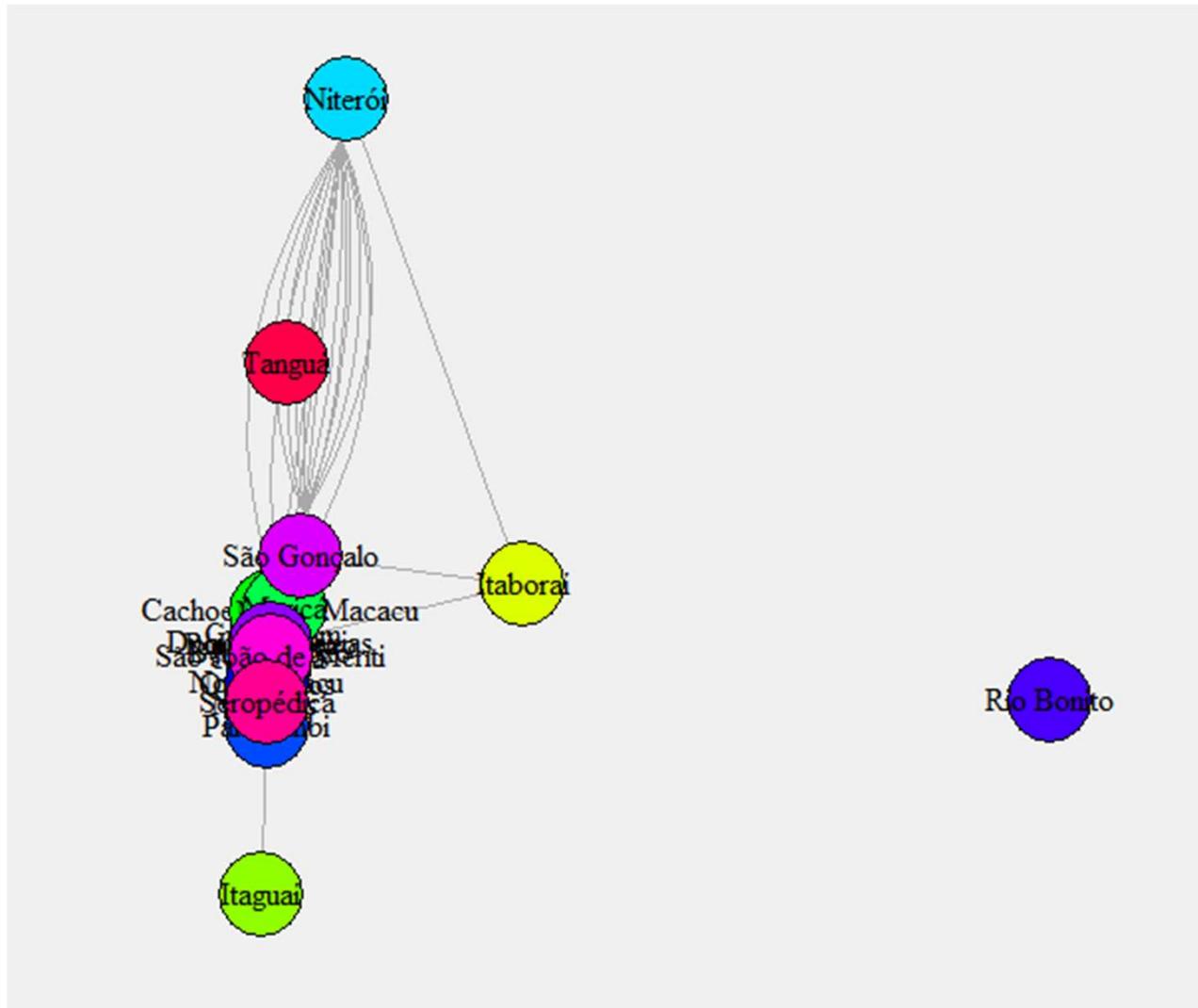
Representação espacial da movimentação de **chegada**



Representação espacial da movimentação de **saída** de trabalhadores dos municípios da RMRJ, 2010 (Mattos, 2016).



Representação espacial da movimentação de **chegada** de trabalhadores dos municípios da RMRJ, 2010 (Mattos, 2016).



4. Modelo de intensidade de conexões de abertura e de receptividade

$$Y_{ij} \sim \text{Poisson}(\mu_{ij}), i \neq j$$

$$\mu_{ij} = e_{ij}\lambda_{ij}$$

$$\log(\lambda_{ij}) = c_0 + b_i + a_j$$

onde e_{ij} é um offset representando o número esperado de conexões entre os elementos i e j da rede caso todos os elementos tivessem o mesmo padrão de conexões.

Interpretação das componentes

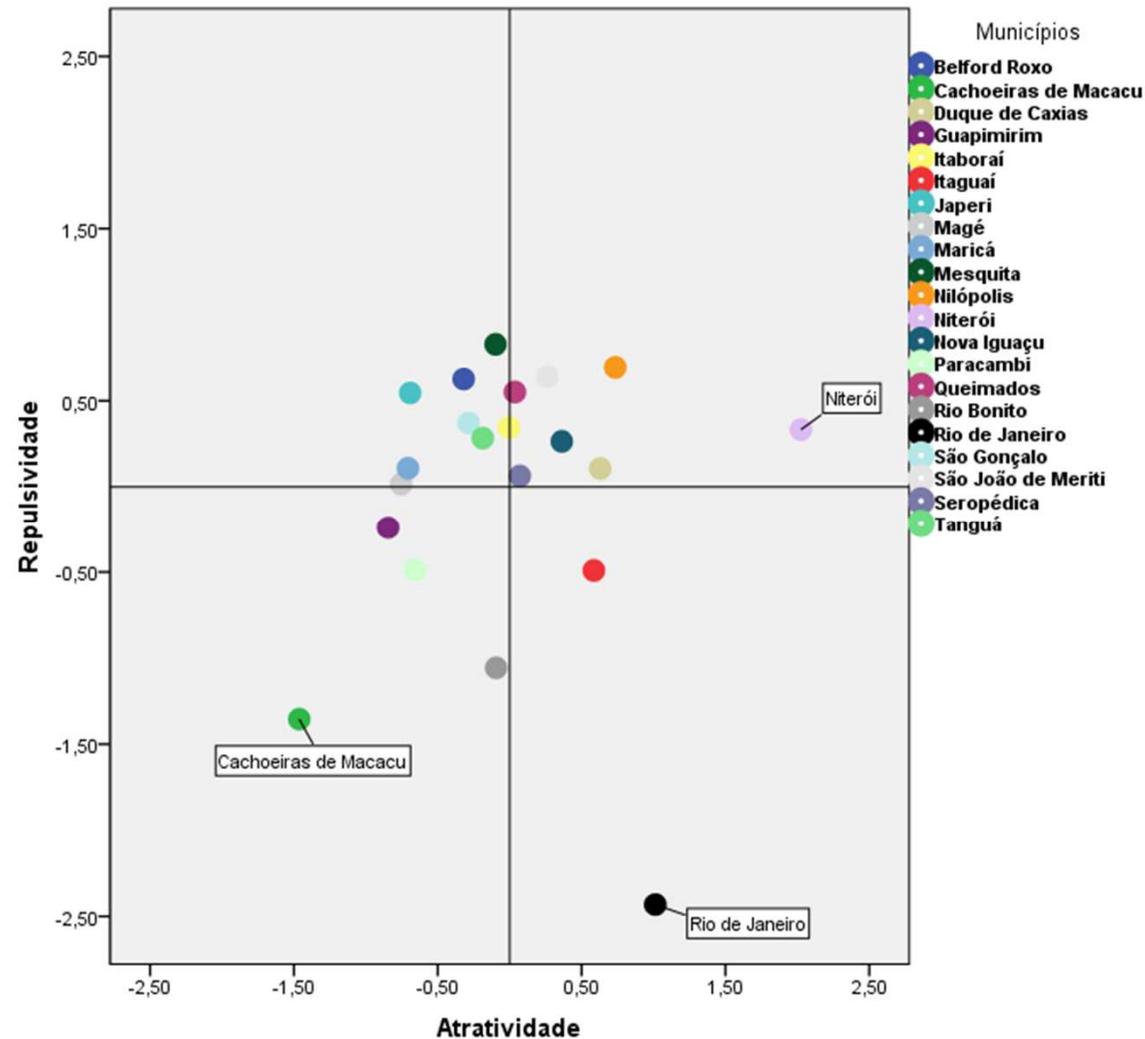
- ▶ b_i representa o grau de abertura para novas conexões do i -ésimo elemento da rede;
- ▶ a_j representa o grau de receptividade para novas conexões do j -ésimo elemento da rede;
- ▶ Quanto maior o valor destes efeitos, **maior** será o número esperado de conexões entre os elementos i e j .

Pendularidade na Região Metropolinada do Rio de Janeiro

No caso da utilização deste modelo para dados de fluxos pendulares na RMRJ, temos:

- ▶ b_i representa o índice de repulsividade do i -ésimo município;
- ▶ a_j representa o índice de atratividade do j -ésimo município;
- ▶ e_{ij} é o fluxo esperado entre os municípios i e j caso toda a Região Metropolitana tivesse o mesmo fluxo pendular.

Distribuição dos municípios da RMRJ segundo índices de atratividade e repulsividade para movimento pendular, 2010 (Mattos, 2016).



Considerações Finais

- ▶ A aplicação de modelos para redes sociais é eficaz para organizar e identificar padrões e agrupamentos de fluxos migratórios e pendulares;
- ▶ É possível interpretar as distâncias das posições de cada elemento da rede e ainda estimar a probabilidade de aumentos ou diminuições nos fluxos entre dois elementos da rede;
- ▶ As precisões das estimativas podem ser obtidas com algum trabalho computacional;

Considerações Finais

- ▶ Os resultados obtidos com a aplicação destes modelos permite a obtenção de **índices** de **atratividade** e **repulsividade** de localidades, levando-se em conta todo o padrão da rede;
- ▶ Adicionalmente, os resultados permitem uma melhor **visualização** da dinâmica dos fluxos migratórios.

OBRIGADO!

gustavo.ferreira@ibge.gov.br