Projeto URBISAmazônia

Relatório final de bolsa

Relatório apresentado à Fundação de Ciência, Aplicações e Tecnologias – FUNCATE - relativo à concessão de bolsa de pesquisa na categoria Desenvolvimento Tecnológico e Industrial I-A

Período: 01/06/2013 a 31/09/2013 (4 meses)

Bolsista: Flávia da Fonseca Feitosa

Assinatura do bolsista

Assinatura do coordenador do projeto no INPE:

Dr. Antonio Miguel Vieira Monteiro

Abril de 2014

População, Espaço e Ambiente na Amazônia:

Redes Migratórias e Estruturação Urbana no Estado do Pará.

SUMÁRIO

População, Espaço e Ambiente na Amazônia:

Redes Migratórias e Estruturação Urbana no Estado do Pará.

1. APRESENTAÇÃO

Este projeto, com duração inicial prevista de 12 meses, teve como objetivo explorar a utilização de técnicas de análise de redes complexas espacialmente explícitas para o estudo do papel dos fluxos migratórios na estruturação do espaço urbano-regional no Estado do Pará. O projeto visava contribuir para o aprimoramento da integração entre distintos núcleos focais do projeto Urbis Amazônia: o núcleo da demografia, com sede no NEPO/Unicamp e o núcleo de geotecnologias para estudos urbanos, com sede no INPE.

O projeto foi, no entanto, interrompido em virtude da aprovação da bolsista em concurso público para professor adjunto na Universidade Federal do ABC (UFABC), que assumiu a nova posição a partir de outubro de 2013.

Durante o período de bolsa, a bolsista esteve envolvida com as seguintes atividades:

a. Participação na organização da reunião "meso escala" do projeto URBIS, realizada no dia 10 de julho de 2013 no INPE, com o objetivo de promover uma revisão do escopo teórico-conceitual para integração dos trabalhos MESO e MICRO escalas. A reunião contou com a presença de pesquisadores e bolsistas do INPE, ITV DS, FACE-UFMG, NEPO-Unicamp e FGV.

b. Participação na organização do II Seminário Nacional sobre População, Espaço e Ambiente - (Re)Pensando o campo de População, Espaço e Ambiente: Experiências/Tendências Contemporâneas e Perspectivas Futuras - da Associação Brasileira de Estudos Populacionais (ABEP). Este encontro foi realizado no INPE, nos dias 29 e 30 de outubro de 2013, e teve um papel estratégico na consolidação do diálogo entre áreas relevantes do projeto Urbis Amazônia vinculadas à demografia e geotecnologias.

c. Revisão bibliográfica preliminar sobre técnicas de análise de redes complexas, com potencial para a análise das relações entre as dinâmicas demográficas na Amazônia, particularmente a estrutura da rede migratória, e suas relações com a configuração do espaço urbano-regional no Pará. A seguir, uma descrição sucinta deste levantamento é apresentada.

2. TÉCNICAS DE ANÁLISE DE REDES: CONCEITOS-CHAVE E MEDIDAS

O conceito de rede é definido por um conjunto de itens, vértices ou nós, que se conectam entre si por meio de linhas, os arcos (NEWMAN, 2003). O estudo de redes é bastante antigo e tem muita tradição na matemática e na sociologia.

Na matemática, baseia-se na teoria dos grafos: a solução de Euler para o problema da ponde Königsberg em 1758 é considerada a primeira prova na teoria das redes. Já na sociologia, destaca-se o estudo de redes sociais, no qual modela-se os relacionamentos de um determinado grupo e analisa-se os padrões de comportamento de seus componentes. Estas análises se dão por meio de métricas topológicas das redes que dentre outras coisas indicam a centralidade de um indivíduo na rede, e a existência de comunidades dentro da rede.

Além da matemática e da sociologia, outras ciências também se dedicam ao estudo das redes. A biologia (redes de moléculas), a ecologia (redes de cadeias alimentares), a geografia (redes de transporte, redes de cidades, redes de mobilidade, e redes hidrográficas - Figura 1), ciência da computação (web e internet) entre outras aplicações.

O maior movimento hoje em pesquisa de redes se dá no estudo de propriedades estatísticas e modelos de desenvolvimento de redes de larga escala. Isso ocorre devido à disponibilidade de computadores e de redes de comunicação que permitem a coleta e análise de dados em grande escala. Para a análise de redes muito extensas, foi necessário o desenvolvimento de novas medidas estatísticas, conforme será apresentado na seção 2.2. Na próxima seção serão apresentados conceitos fundamentais ao entendimento destas medidas, bem como a classificação de alguns tipos de redes.

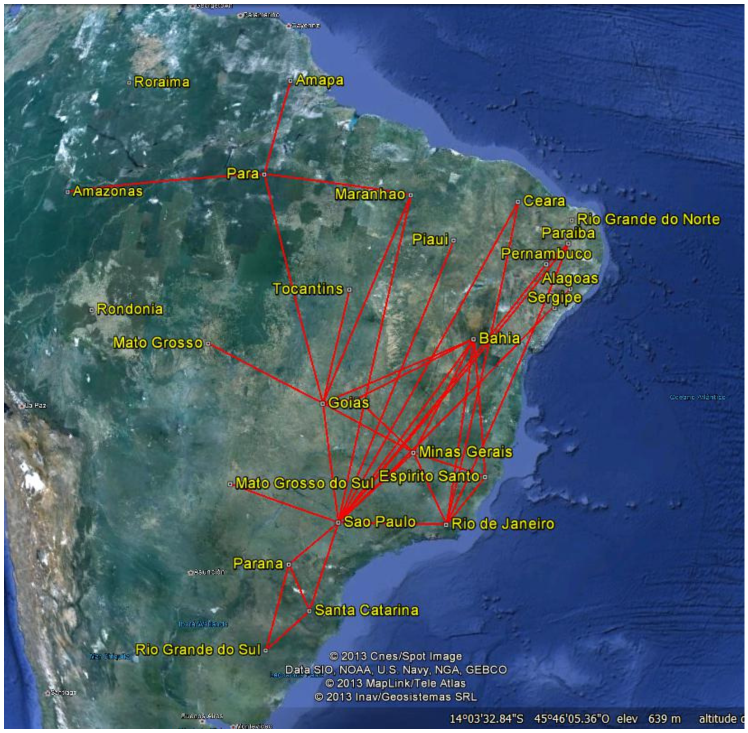

Figura 1 - Exemplo de rede espacial a partir de dados de migração recente. 2010.
Fonte: Censo demográfico 2010

2.1 Conceitos fundamentais

Entre os conceitos fundamentais para a compreensão de uma rede, podemos citar os seguintes:

a) Vértice: unidade fundamental da rede, também denominada como *nodo*, *site* (na Física), *nó* (na Ciência da Computação) ou *ator* (Sociologia).
b) Aresta ou Arco: é a linha que conecta dois vértices. Também conhecida como *bond* (na Física, significa ligação ente átomos), *link* (Ciência da Computação) e *tie* (na Sociologia, significa laço, vínculo).
c) Vértices Adjacentes: são vértices que são ligados por um arco.
d) Número de Vizinhos de um Vértice: são todos os vértices adjacentes ao vértice em questão. Em uma rede não orientada o número de vizinhos de um vértice é a sua valência.
e) Arco Orientado/Não-Orientado: um arco é considerado orientado quando ele possui apenas um sentido, como por exemplo, uma rua de mão única. E ele é não orientado quando possui dois sentidos de conexão. O sentido das conexões é representado por setas.

f) Grau ou Valência: O número de arcos conectados a um vértice. O grau não é igual ao número de vértices adjacentes a um vértice, pois é possível que exista mais de um arco conectando dois vértices. Em um gráfico orientado há o grau de entrada (número de arestas que chegam a um vértice) e grau de saída (número de arcos que saem de um vértice). Em análise de redes sociais também é conhecido com grau de centralidade.

g) Componente: É o conjunto de vértices que podem ser alcançados por um determinado vértice, passando por arcos que conectam este conjunto. Na rede orientada um vértice tem um componente de entrada e um de saída.

h) Componente Gigante: É o maior componente de rede (em número de arcos).

i) Caminho (Path): É uma rota entre dois vértices da rede sem a repetição de vértices, com exceção do vértice inicial e final.

j) Trilha (Trail): É uma rota entre dois vértices da rede sem a repetição de arcos, com exceção do vértice inicial e final.

k) Itinerário (Walk): É uma rota entre dois vértices da rede sem restrições.

l) Distância Geodésica ou Caminho mais Curto: É o caminho mais curto pela rede que liga um vértice a outro (medido em número de arcos).

m) Matriz de Vizinhança: É uma matriz que representa as relações de vizinhança entre os vértices da rede. Quando um vértice se liga a outro, a célula da matriz recebe valor 1, e quando não há ligação recebe o valor zero.

n) Matriz Geodésica: É uma matriz que reúne o caminho geodésico de todos os pares de nós da rede.

o) Matriz de Proximidade: É a matriz com a distância euclidiana entre todos os pares de vértices da rede.

p) Diâmetro da Rede: É o comprimento (novamente em número de arcos) do caminho geodésico mais comprido da rede.

q) Tamanho da Rede: É o número total de vértices.

2.2 Tipos de redes

Há redes de diferentes tipos, segundo a combinação variada de tipos de vértices e arcos e restrições de relacionamento. Uma rede que representa uma determinada cadeia alimentar pode ser representada, por exemplo, por dois conjuntos de vértices: os predadores e as presas. Os arcos também podem representar propriedades distintas, como intensidade do fluxo, direção ou tipo de relacionamento. Entre as combinações possíveis de características de nós, arcos e relacionamentos podem-se destacar os seguintes tipos de rede:

a) Rede orientada ou dirigida (também conhecida como um grafo do tipo diágrafo): Onde os fluxos seguem um sentido específico, o exemplo da cadeia alimentar é uma boa representante deste tipo. A rede orientada pode ser de dois tipos, cíclica e acíclica, o primeiro se dá quando as interações entre os vértices ocorrem em loops, e o segundo quando não há este tipo de interação.
b) Redes bipartidas: Redes em que só se estabelecem conexões entre nós de tipos diferentes.
c) Redes planares: São redes nas quais os arcos não se cruzam. Um bom exemplo são as redes de transporte rodoviário.

2.3 Medidas em redes

Nesta seção são apresentadas as principais medidas das redes com ênfase nas medidas baseadas na topologia das redes. Esta escolha foi baseada em dois fatores: (a) são as métricas mais aplicadas nos estudos recentes que aplicam teoria; (b) são básicas ao entendimento de métricas mais complexas que integram fluxos e atributos espaciais.

2.3.1 Densidade da rede

É a razão entre o número de arcos existentes e o número total de arcos possíveis. Esta medida fornece uma noção de quanto uma rede está próxima ao ponto de ser totalmente conectada. Para redes não orientadas é definida por:

$$d = \frac{a}{n(n-1)}, \tag{1}$$

Onde, a é o número total de arcos na rede e n o número total de vértices.

Para redes orientadas é definida pela seguinte expressão:

$$d = \frac{a}{\frac{n(n-1)}{2}}, \tag{2}$$

2.3.2 Medidas de centralidade

As medidas de centralidade são utilizadas geralmente na análise de redes sociais para identificar o quanto um ator da rede é importante (central) para a comunicação da mesma. As principais medidas de centralidade são: o grau de centralidade, o grau de proximidade (*closeness*) e o grau de intermediação (*betweeness*). O grau de centralidade é o grau ou valência do vértice, o número de arcos conectados a ele. Há uma versão normalizada desta medida na qual a valência do vértice é dividida pelo número de nós da rede menos uma unidade. Esta medida indica o potencial de comunicação de um ator. Assim, quanto mais arcos se conectam a um vértice, maior o seu poder de difundir uma informação.

O grau de proximidade é a soma das distâncias geodésicas de um vértice *i* aos outros vértices da rede. Quando esta distância é pequena, indica que o vértice em questão necessita de distâncias curtas para se conectar com os demais elementos da rede. Assim, esta medida indica o potencial de um vértice se comunicar mais rapidamente com os outros. Como um vértice que possui uma proximidade maior aos demais é associado a um valor numérico maior, a definição matemática desta medida é dada pelo inverso da soma das distâncias geodésicas.

$$c_p = \frac{1}{\sum_j d_{ij}}, \tag{3}$$

Onde d_{ij} é a distância geodésica do vértice *i* ao vértice *j*.

O grau de intermediação de um vértice i mede o quanto este vértice se encontra “no meio do caminho” dos outros vértices da rede. É a razão entre o número de distâncias geodésicas entre dois vértices, *a* e *b*, que passam por *i* e o total de distâncias entre a e b. O grau de intermediação indica o poder de controle do vértice *i* sobre a informação. É expresso por:

$$c_i = \sum \frac{d_{ab(i)}}{d_{ab}}, \tag{4}$$

Onde d_{ab} é o número de distâncias geodésicas entre os vértices a e b e $d_{ab(i)}$ são as distâncias que incluem o vértice i.

2.3.3 Distância geodésica média

É a média das distâncias geodésicas da rede. Esta estatística nos fornece uma análise sobre quão conectado são os vértices da rede, quanto mais próximo de 1 for esta medida maior, é a probabilidade de todos os vértices estarem conectados.

Considerando uma rede não orientada, a distancia geodésica média entre dois pares de vértices na rede é dada por:

$$\ell = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}, \tag{5}$$

Onde, n é o número de nós da rede, d_{ij} é a distância geodésica entre os vértices i e j. O denominador da equação é o maior número possível de arestas. Assim ℓ é nada mais do que a soma de todas as distâncias geodésicas dividida pelo número de arestas, uma média aritmética.

A definição de ℓ (5) não é adequada para redes com mais de um componente com a rede. Neste tipo de rede há par de vértices em componentes diferentes, assim estes vértices não possuem distância geodésica. Neste caso, convencionou-se que a distância geodésica entre estes pares de vértices seria infinita. Entretanto, isto resulta em ℓ infinito também, impossibilitando seu cálculo. Há duas soluções para este problema. A primeira exclui do cálculo pares de vértices localizados em componentes diferentes. A segunda define ℓ como a "média harmônica" da distância geodésica:

$$\ell^{-1} = \frac{1}{\frac{1}{2}n(n+1)} \sum_{i \geq j} d_{ij}^{-1}, \tag{6}$$

Usando esta definição, os valores infinitos de d_{ij} não contribuirão em nada para a soma. Segundo NEWMAN (2003), esta definição é pouco utilizada nos trabalhos que realizam esta medida, no entanto, ela deveria ser mais utilizada, pois evita o trabalho de exclusão de pares de vértices em componentes diferentes.

2.3.4 Transitividade ou coeficiente de *clustering*

O coeficiente de *clustering* é a propensão de dois vértices que possuem um vizinho em comum estarem conectados um ao outro. (NEWMAN e PARK, 2003). Assim se um vértice A está conectado a um vértice B e este último está conectado a um vértice C, há grandes chances de A e C estarem conectados um ao outro. Como diz NEWMAN (2003), em termo de redes sociais: "um amigo do seu amigo, provavelmente é seu amigo também". Topologicamente esta medida é traduzida como o alto número de triângulos na rede, ou seja, conjunto de três vértices que são conectados entre si. A expressão pode ser definida por:

$$C = \frac{3 \times número\ de\ triângulos}{número\ de\ triplas\ conectadas}, \tag{7}$$

Uma tripla conectada é um vértice conectado a outros dois vértices sem a formação de um triângulo. O numerador da expressão é multiplicado por três porque cada triângulo contribui três vezes para a formação de uma tripla. Desta forma, também é garantido que os valores de C estejam no intervalo entre 0 e 1.

WATTS et al. (1998) propuseram o cálculo de um coeficiente local, que é a razão entre o número de triângulos e de triplas de cada vértice i. Este coeficiente local é bastante utilizado na literatura de sociologia onde é conhecido como "densidade da rede". A expressão é definida por:

$$C_i = \frac{número\ de\ triângulos_i}{número\ de\ triplas\ conectadas_i}, \tag{8}$$

O cálculo do coeficiente de *clustering* para toda a rede é a média de todos os coeficientes. Para os vértices de graus 1 ou 0 que possuem numeradores e denominados iguais a 0 foi definido que $C_i = 0$. A expressão para o coeficiente C para toda a rede é definida por:

$$C = \frac{1}{n}\sum_{i} C_i \quad , \tag{9}$$

2.4 Weighted networks: Redes com representação de intensidade de seus fluxos

As seções anteriores deste capítulo se concentraram na apresentação de conceitos básicos da teoria dos grafos e na descrição das métricas mais conhecidas que descreve a topologia e a estrutura das redes. É claro que a compreensão da estrutura de uma rede é fundamental para o entendimento do funcionamento do sistema que esta representa, mas os fluxos que circulam sobre esta rede também o são. As redes do mundo real, entre elas as de migração, além de possuírem estruturas topológicas complexas, também contém diferentes capacidades e intensidades de conexões em seus arcos. A representação e quantificação dos fluxos das redes se mostram relevantes em diversas aplicações: em redes de cadeia alimentar, na web, em redes sociais, em redes de transporte, etc. (BARTHE´LEMY et al., 2005).

Neste contexto, é reconhecida a necessidade de instrumentos analíticos para a análise dos fluxos das redes. BARTHE´LEMY et al. (2005) contribuem neste sentido, propondo algumas medidas que relacionam a topologia das redes com seus fluxos. Além disso, os mesmos autores também criaram um modelo simples de evolução de redes que leva em consideração seus fluxos. No mesmo trabalho foi avaliada a efetividade das medidas em dois conjuntos de dados, rede de transporte aéreo, e rede de colaboração científica, e os autores conseguiram alcançar bons resultados, suas medidas refletiam as características das redes e revelaram traços bem interessantes a respeito dos fluxos das mesmas.

O conjunto de ferramentas proposto por BARTHE´LEMY et al. (2005) foi também avaliado em outro experimento. As medidas foram aplicadas em uma análise sobre o movimento de migração pendular (*commuting*) de municípios da região de Sardina na Itália (MONTIS (2007)). O mesmo conjunto de ferramentas foi conjugada com algumas medidas espaciais para elaborar um modelo de rede para a representação e compreensão da evolução de uma rede de tráfego aéreo BARRAT et al.(2005).

Algumas das medidas propostas por BARTHE´LEMY et al (2005) são:

- A matriz de pesos, na qual o valor do fluxo é representado entre os pares de vértices em uma matriz quadrada;
- Caracterização da distribuição dos pesos da rede. Enquanto na caracterização topológica se analisa gráficos da probabilidade que um vértice tenha um grau k. Similarmente, a primeira caracterização dos fluxos (pesos) é obtida pela distribuição P(w) que um dado arco tenha um determinado peso w;
- Força é uma medida variante do grau de um vértice. É o total de fluxos de um vértice;
- E disparidade que fornece uma visão do quão homogêneos ou não são os pesos dos fluxos que se direcionam para um determinado grau. Indica se há a presença de algum arco com peso dispare em relação aos outros arcos vizinhos. A disparidade é expressa por:

$$y(i) = \sum_{j \in v(i)} [\frac{w_{ij}}{s_i}]^2 , \tag{10}$$

Onde, w_{ij} é o peso de um arco que liga o vértice *i* ao vértice *j* , s_i é a força do vértice *i*.

Além dos trabalhos recentes que agregam medidas de peso às medidas topológicas, há uma grande tradição de estudos de fluxos de redes que empregam os modelos de interação espaciais. Interação espacial é definida como o movimento de pessoas, commodities, capital e informação sobre o espaço que resultam de um processo de decisão (HAYNES e FOTHERINGHAM, 1984). Neste modelo o individuo envolvido em uma interação analisava de alguma forma se o benefício da interação superava os custos que eram necessários para superar a distância geográfica entre ele e o seu destino. (FISHER, 2006).

Existem vários modelos de interação espacial, os mais populares são os modelos gravitacionais baseados na lei de gravidade de Newton. Os modelos gravitacionais foram largamente usados em análises de diversos fluxos como migração, comércio, transporte, informação, etc. A forma mais geral dos modelos gravitacionais mensura um fluxo por meio do produto das massas dos dois vértices, multiplicado pelo inverso da distância entre eles HAGGETT (1971). A formula é dada por:

$$M_{ij} = P_i P_j \times \frac{1}{d_{ij}}, \quad (11)$$

A massa do modelo poderia ser qualquer variável que expressasse a massa do sistema, como, no caso deste trabalho, população. E a distância poderia ser a distância Euclidiana, a distância geodésica em uma rede, o custo ou tempo.

Inúmeros modelos foram criados a partir de modificações dos modelos gravitacionais, inserindo variáveis como a impedância, por exemplo. A impedância é caracterizada como sendo uma medida que avalia os efeitos contrários à realização ou ao prolongamento das viagens. Em uma rede de transportes pode-se modelar as condições das estradas, ou a possibilidade de engarrafamentos como a impedância. Além dos modelos gravitacionais também há modelos de decaimento da distancia, modelos de maximização da entropia, modelos baseados em lógica *fuzzy*, modelos que utilizam redes neurais artificiais, entre outros.

A maioria dos modelos citados acima apenas faz estimativa dos fluxos de trocas entre os vértices. Eles raramente relacionam os fluxos com a topologia das redes.

Além de incorporar a análise dos fluxos, as atuais análises sobre a estrutura das redes deveriam incorporam também análises sobre os atributos espaciais das mesmas. Há poucos trabalhos que se dedicam a incorporar a dimensão espacial. Um deles é o de GASTNER e NEWMAN (2006) que entre outras medidas analisam a distribuição do tamanho real dos arcos de três redes geográficas, uma de estradas, a outras de tráfego aéreo e da internet, e verificaram que na rede de estradas o comprimento médio era bem menor do que na rede de tráfego aéreo. Neste mesmo trabalho os autores criaram um modelo de crescimento de redes que obedece a duas restrições básicas: os vértices da rede devem se conectar com as menores distâncias euclidianas possíveis, e com os menores arcos possíveis também. Com estas restrições eles criaram redes que se aproximam a estrutura das redes espaciais criadas.

3. CONSIDERAÇÕES FINAIS

Os recentes avanços no estudo da estrutura das redes fornecem ferramentas analíticas importantes para a compreensão dos fenômenos que se dão nos espaços articulados por elas. No caso deste trabalho, no espaço urbano-regional no Estado do Pará. Dada a escassez de trabalhos que explorem a análise dos fluxos sob uma perspectiva que incorpore os atributos espaciais da rede, a intenção da revisão de literatura conduzida neste período de bolsa é a de subsidiar futuras análises que integrem estrutura, fluxos migratório e atributos espaciais.

4. REFERÊNCIAS BIBLIOGRÁFICAS

BARRAT, A., M. BARTH´ELEMY e A. VESPIGNANI. The effects of spatial constraints on the evolution of weighted complex networks. Journal of Statistical Mechanics: Theory and Experiment, n.5, p.20. 2005.

BARTHE´LEMY, M., A. BARRAT, R. PASTOR-SATORRAS e A. VESPIGNANI. Characterization and modeling of weighted networks. Physica A, v.346, p.12. 2005.

BATTY, M. Network Geography: Relations, Interactions, Scaling and Spatial Processes in GIS. CASA Working Papers. 63: 24 p. 2003.

FISHER. Spatial Interaction Models and the Role of Geographic Information Systems. In: FISHER (Ed.). Spatial Analysis and GeoComputation Selected Essays Berlin: Springer Berlin Heidelberg, 2006. Spatial Interaction Models and the Role of Geographic Information Systems, p.29-43

GASTNER, M. T. e M. E. J. NEWMAN. The spatial structure of networks The European physical journal. B, Condensed matter physics, v.49, n.2, p.247. 2006.

HAGGETT, P. Locational Analysis in Humann Geography. London: Edward Arnold 1971. 339 p.

HAYNES e FOTHERINGHAM. Gravity and spatial interaction models. . Beverly Hills: SAGE. 1984

MONTIS, A. D., M. BARTH´ELEMY, A. CHESSA e A. VESPIGNANI. The structure of Inter-Urban traffic A weighted network analysis. Environment and Planning B: Planning and Design, v.34, n.5, p.19. 2007.

NEWMAN, M. E. J. The structure and function of complex networks. published in SIAM, v.45, p.89. 2003.

NEWMAN, M. E. J. e J. PARK. Why social networks are different from other types of networks. Phys. Rev. E, v.68. 2003.

PINHO, C.M.D. Análise das redes de localidades ribeirinhas amazônias no tecido urbano estendido: uma contribuição metodológica. Tese de Doutorado. Programa de Pós-Graduação em Sensoriamento Remoto, Instituto Nacional de Pesquisas Espaciais. São José dos Campos, 2012.

WATTS, D. J., D. J., STROGATZ e S. H. Collective dynamics of 'small-world' networks. Nature Publishing, v.393, n.6684, p.42. 1998.