Navegação ótima em redes complexas

D. O. Cajueiro

Departamento de Economia (UNB) e INCT (Sistemas Complexos)

2009

Daniel O. Cajueiro
Quem é?

- 1995-1998: Iniciação Científica em Física no Projeto Fenômenos Complexos em Catalisadores (UFBA) $\Rightarrow$ Fractais e Modelos de Crescimento de superfícies rugosas.
- 1998: Graduado em Engenharia Química (UFBA)
- 2000: Mestre em Engenharia Eletrônica e Computação – Sistemas e Controle (ITA) $\Rightarrow$ Inteligência Artificial, Controle Adaptativo e Estabilidade de Sistemas Não Lineares.
- 2002: Doutor em Engenharia Eletrônica e Computação – Sistemas e Controle (ITA) $\Rightarrow$ Otimização Estocástica em tempo contínuo, Finanças, Pesquisa Operacional e Aprendizado por Reforço.
- Atual: Professor de Matemática, Econometria, Estatística, Finanças e Microeconomia do Departamento de Economia (UNB) e Membro do INCT (Sistemas Complexos).

Ir de um nó a outro utilizando as arestas dessa rede.
Exemplos:

- Deslocamento de um andarilho em uma cidade
- Transmissão de sinais entre dois computadores
- Busca de uma página na WWW
- Busca de um determinado indivíduo dentro uma organização

- Identificar a habilidade de comunicação entre dois indivíduos em uma rede
- Investigar como a estrutura de uma rede restringe a comunicação (ou o movimento) entre os nós dessa rede
- Criar heurísticas ótimas/sub-ótimas para realizar busca de nós em redes reais
- Projetar redes com topologias ótimas no contexto de busca local

Se divide principalmente em três ramos:

- Navegação de andarilhos aleatórios ("random walkers")
- Navegação de andarilhos direcionados ("directed walkers")
- Busca em redes complexas

Um random walker localizado em determinado nó da rede escolhe um vizinho baseado em alguma matriz de transição.

Contribuições:

- Cálculo do tempo médio de primeira passagem ("mean first passage time") para redes arbitrárias e introdução do "random walk centrality" [Noh and Rieger. PRL 92, 118701 (2004)]
- Estudo de várias especificações diferentes para a matriz de transição [Yang. PRE 71,016102 (2005)]
- Amostragem de redes complexas utilizando random walkers e comparação das propriedades da rede amostrada com a rede real [Costa (INCT-SC) e Travieso. PRE 75, 016102 (2007)]

Um directed walker vai sempre de um nó a outro através do caminho mais curto. Ele é como um andarilho em uma nova cidade que perguntando a pessoas na rua, chega ao seu destino através do caminho mais curto.

- Introdução do "search information" que é a informação total necessária para ir-se de um nó a outro em uma rede $\Rightarrow$ "node with minimal access information" and "the best node to be hidden" [Sneppen *et al.*. EL 69, 853 (2005). Rosvall *et al.* PRL 94, 028701 (2005). Rosvall *et al.* PRE 72, 046117 (2005)

- Suponha que nós temos uma cidade representada por uma rede G com n nós $V(G) = \{1, 2, \cdots, t, \cdots, n\}$ onde t é um nó especial chamado alvo.
- Em cada nó, um viajante deve escolher entre realizar o próximo passo aleatoriamente a um custo C_N ou obter informação sobre o caminho correto que deve seguir e realizar o próximo passo com custo $C_N + C_I$, onde C_N é o custo de navegação e C_I é o custo de informação.

- Assume-se que o viajante em cada nó toma a decisão com o objeto de minimizar o custo da trajetória dada por

$$J(i) = \min_{\pi \in \Pi} E_\pi \left[\sum_{k \in \mathcal{P}(i,t)} g(k, u(k))/i \right]$$

onde $\mathcal{P}(i,t)$ é o caminho do nó i ao alvo t, a expectativa $E^\pi[\cdot/i]$ é condicional a política π e ao nó i, $\pi = \{u(1), u(2), \cdots, u(t), \cdots, u(n)\}$ é uma política admissível pertencente ao conjunto de políticas admissíveis Π e $u(k)$ é um controle admissível pertencente ao conjunto de controles admissíveis $U(k) = \{0, 1\}$, $\forall k$.
- Assume-se que $u(k) = 0$, se no nó k ele decide escolher aleatoriamente o próximo passo e $u(k) = 1$, se no nó k ele decide perguntar a outras pessoas que direção deve seguir.

- O custo por estágio $g(k, u(k))$ é o custo de usar $u(k)$ no nó k dado por

$$g(k, u(k)) = \begin{cases} C_N, \text{if } u(k) = 0, \forall k \neq t \\ C_N + C_I, \text{if } u(k) = 1, \forall k \neq t \end{cases}$$

Crítica: Na prática o viajante em uma nova cidade nunca poderá resolver exatamente esse problema, pois ele precisaria conhecer a priori a estrutura da rede da cidade dada por $p_{kl}(u), \forall u \in U(k)$ e $k, l \in V(G)$ – a matriz de transição de um nó k para um nó l se a lei de controle u é usada no nó k.

Entretanto, vale a pena resolver esse problema pois:

- É um limitante superior para a qualidade do comportamento do viajante.
- É interessante para caracterizar redes complexas.
- Extrapola os comportamentos dos random walkers e directed walkers. Se C_I/C_N é grande o suficiente, nunca será ótimo perguntar pela direção correta. Se C_I/C_N é baixo, sempre será ótimo perguntar sobre a direção correta.

- Uma vez que $U(k)$ é finito para todo k e assumindo que a rede não clusters isolados, $p_{tt}(u) = 1$ para todo $u \in U$ e $g(t,u) = 0$ para todo $u \in U$, esse problema pode ser formulado por um problema de caminho mais curto estocástico e a solução é dada pela equação de Bellman

$$J(i) = \min_{u \in U(i)} \left[g(i,u) + \sum_{j=1}^{n} p_{ij} J(j) \right]$$

- Uma vez que $g(t,u) = 0$ e $p(t,j) = 0, \forall j \neq t$, então $J(t) = 0$.
- Essa equação diz que $J(i)$, o custo ótimo que o viajante no nó $i \in V(G)$ tem que pagar para alcançar o nó t, é dividido em duas partes: o custo que o viajante tem que pagar no nó i mais o custo que o viajante tem que pagar em todos os outros nós antes de alcançar t.

- Defina $\overline{J}^N(i)$, $\forall i \in V(G)$ como a média de $J(i)$ para todos os alvos t.
- Se $\pi = \{1, 1, \cdots, 1\}$ então $\overline{J}^N(i)$ é o comprimento característico do nó multiplicado pelo custo $C_N + C_I$.
- Se $\pi = \{0, 0, \cdots, 0\}$, $\overline{J}^N(i)$ é o tempo médio de primeira passagem por t a partir de i multiplicado por C_N.
- Portanto, $\overline{J}^N(i)$ mede em média a dificuldade de navegar do nó $i \in V(G)$ para qualquer outro nó $j \in V(G) - \{i\}$, se o viajante usa a política ótima.
- Essa variável estende a idéia da chamada *minimal access information*.

- Defina $\overline{J}^{H}(i)$, $\forall i \in V(G)$ como a média de todos $J(j), \forall j \in V(G) - \{i\}$, quando $t = i$. Ela mede a habilidade do nó ficar escondido na rede generalizando a idéia da chamada *hide*.
- Defina $\overline{\overline{J}}$ como a média de J para todos os nós e todos os alvos. Ela mede a dificuldade total de navegar na rede.

- Aplicamos essas idéias a uma rede livre de escala com 250 nós.
- Dividimos os nós em duas amostras de tamanho 125 de acordo com o tamanho de $\overline{J}^N$ when $C_I/C_N = 0$.
- $\overline{J}^N_{\mathrm{high}}$ ($\overline{J}^N_{\mathrm{low}}$) é dito ser a média de $\overline{J}^N$ para a fração de nós que têm os mais altos (baixos) $\overline{J}^N$ quando $C_I/C_N = 0$.
- Uma vez que essa divisão é feita quando $C_I/C_N = 0$, os nós com mais baixos (altos) $\overline{J}^N$ são exatamente aqueles que têm os menores (maiores) comprimentos característicos.

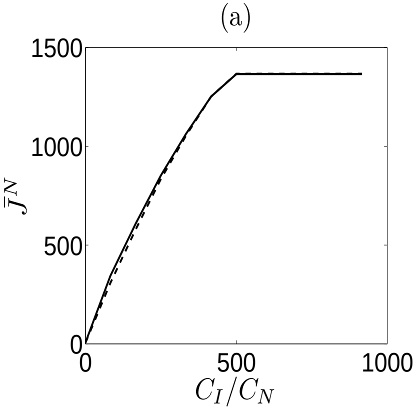

Figura: (a) mostra $\overline{J}^N_{\mathrm{high}}$ (linha sólida) e $\overline{J}^N_{\mathrm{low}}$ (linha tracejada) para vários valores de C_I/C_N.

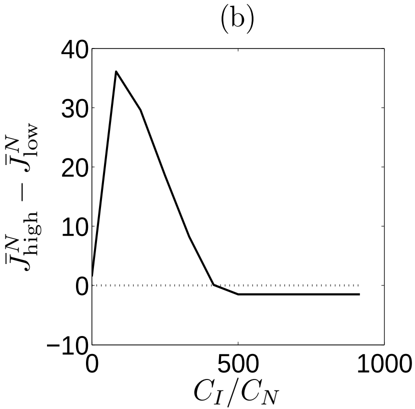

Figura: (b) mostra $\overline{J}^N_{\mathrm{high}} - \overline{J}^N_{\mathrm{low}}$ para vários valores de C_I/C_N.

Navegação ótima em redes livres de escala

Transição de regime

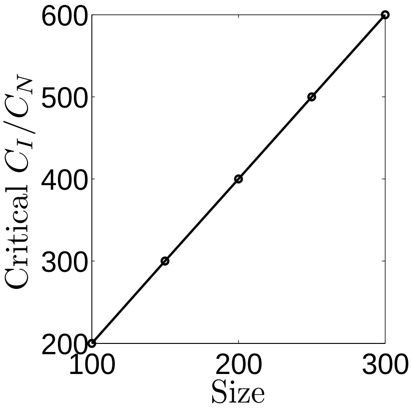

Figura: A relação entre a razão crítica C_I/C_N e o tamanho de uma rede típica livre de escala.

- A transição como apresentada acima ocorre em todas redes em que existir uma correlação positiva entre a amostra de nós usada para construir $\overline{J}^{N}_{\mathrm{high}}$ e os nós com menores graus da rede.
- A linha quase reta na relação entre a razão crítica e o tamanho da rede ocorre devido a conexão preferencial que ocorre nessas redes.

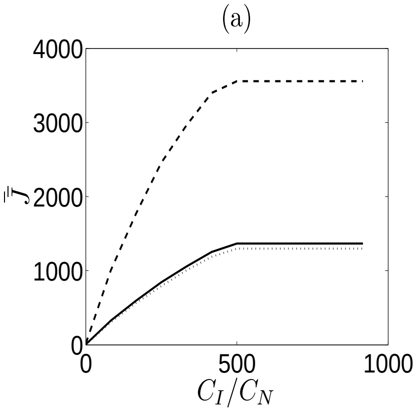

Figura: (a) mostra $\bar{\bar{J}}$ para vários valores de C_I/C_N para uma rede livre de escala (linha sólida), para uma versão dessa rede com máxima hierarquia (linha pontilhada) e para uma versão dessa rede com mínima hierarquia (linha tracejada).

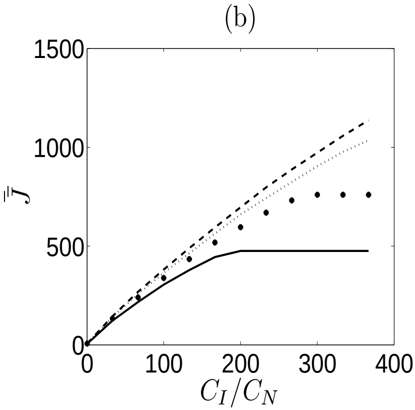

Figura: (b) mostra $\overline{\overline{J}}$ para redes de livres de escala típicas para vários valores de C_I/C_N e para diferentes tamanhos: 100 (linha sólida), 150 (star), 200 (linha pontilhada) and 250 (linha tracejada).

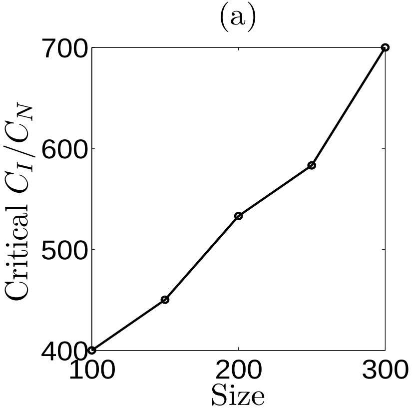

Figura: (a) mostra a relação entre a razão crítica C_I/C_N e o tamanho de uma rede aleatória típica.

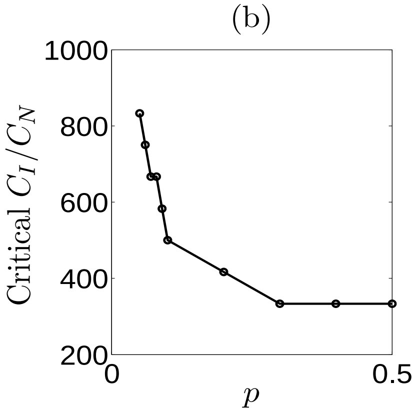

Figura: (b) mostra a relação entre a razão crítica C_I/C_N e a conectividade de uma rede aleatória típica.

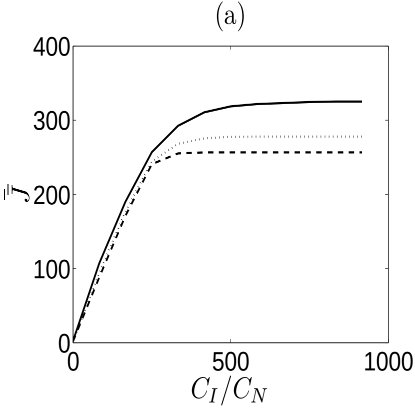

Figura: (a) $\bar{\bar{J}}$ para redes aleatórias típicas para vários valores de C_I/C_N com diferentes probabilidades de conexão: $p = 0.05$ (linha sólida), $p = 0.1$ (linha pontilhada) and $p = 0.2$ (linha tracejada).

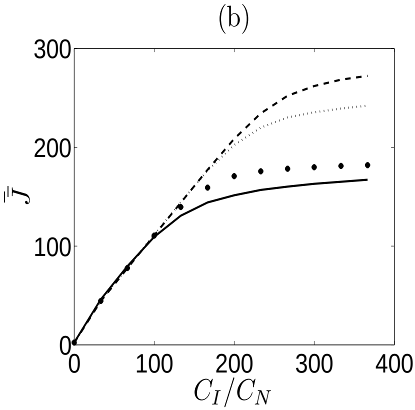

Figura: (b) $\bar{\bar{J}}$ para redes aleatórias típicas para vários valores de C_I/C_N com diferentes tamanhos: 100 (linha sólida), 150 (estrela), 200 (linha pontilhada) and 250 (linha tracejada).

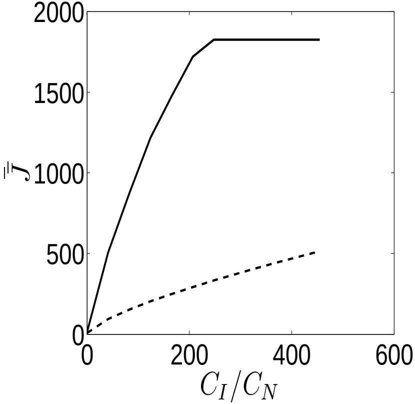

Figura: Essa figura compara a dificuldade de navegação $\overline{\overline{J}}$ no sistema de Metrô de Boston (linha sólida) com a média de 10 simulações de sua versão aleatorizada (linha tracejada).

Navegação ótima em redes reais

Dificuldade de navegação na rede do clube de karatê de Zachary

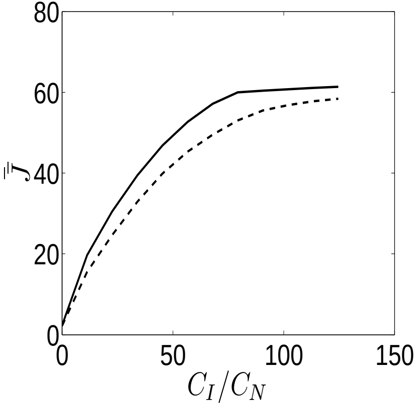

Figura: Essa figura compara a dificuldade de navegação $\overline{\overline{J}}$ na rede do clube de karatê de Zachary (linha sólida) com a média de 10 simulações de sua versão aleatorizada (linha tracejada).

- Em geral, é mais difícil navegar em redes redes reais que em suas contrapartidas aleatórias.
- Isso ocorre devido as restrições reais que essas redes estão sujeitas.
- Ao estudar navegação em redes, não é suficiente olhar para os extremos.

- Esse trabalho apresentou uma metodologia para lidar com navegação ótima em redes complexas.
- A solução desse problema foi conseguida a partir da solução numérica da equação de Bellman. Para as hipóteses consideradas, essa solução sempre existe.
- Utilizando essa solução foi possível generalizar vários conceitos utilizados na literatura sobre navegação.
- Também foi possível caracterizar o ponto crítico da transição de regime.

- Uma limitação desse trabalho é o mal da dimensionalidade comum em problemas de programação dinâmica. Uma solução para esse problema é considerar aproximações da função valor.
- Nesse trabalho utilizamos redes não direcionadas e não ponderadas, mas em essência ele é válido para redes ponderadas sempre e para redes direcionadas se sempre houver um caminho entre o nó de origem o nó alvo.

Agradecimentos

- Martin Rosvall pela base de dados com cidades suecas.
- Vito Lattora pela base de dados com o sistema de metrô de Boston.
- CNPQ pelo suporte financeiro.