Redes Complexas – CPS765
2018/3

Prof. Daniel R. Figueiredo

Primeira Lista de Exercícios

ATENÇÃO! Para ajudar no processo de aprendizagem, explique suas repostas.

Questão 1: Matriz de Adjacência: A matriz de adjacência A de um grafo $G = (V, E)$ codifica toda sua estrutura. Ou seja, $A_{ij} = 1$ se $(i, j) \in E$ e $A_{ij} = 0$ caso contrário. Desta forma, toda e qualquer informação estrutural de G é obtida através de manipulações de A. Responda as perguntas abaixo.

1. Mostre como A pode ser usada para gerar outra matriz $B^{(k)}$ para $k > 0$ que codifica alcançabilidade em exatamente k passos. Ou seja, $B_{ij}^{(k)} = 1$ se existe ao menos um caminho de comprimento exatamente k entre os vértices i e j, e $B_{ij}^{(k)} = 0$ caso contrário. Mostre uma solução matemática (sem usar um algoritmo), ou seja, $B^{(k)} = f(A, k)$ para alguma função f.

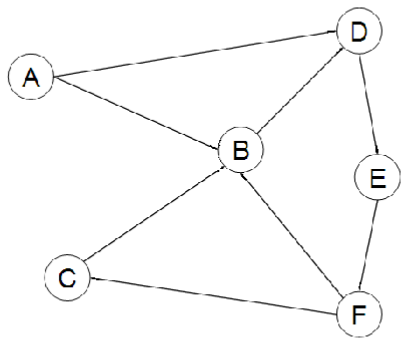
2. Repita o exercício anterior mas agora codificando a alcançabilidade em k ou menos passos. Ou seja, $C_{ij}^{(k)} = 1$ se existe ao menos um caminho de comprimento k ou menor entre os vértices i e j, e $C_{ij}^{(k)} = 0$ caso contrário.

3. Determine a complexidade computacional para calcular $B^{(k)}$ e $C^{(k)}$. Assuma um algoritmo ingênuo para operações algébricas com matrizes.

4. Mostre como reduzir (significativamente) essa complexidade mantendo o algoritmo ingenuo para operações algébricas com matrizes.

Questão 2: Grau médio e densidade: Vimos em aula duas propriedades globais de redes: grau médio e densidade. É possível que uma rede apresente grau médio pequeno e densidade alta, ou vice-versa? Mostre analiticamente a relação entre essas duas métricas, usando apenas n e m (número de vértices e arestas na rede, respectivamente). Atenção para sua definição de "alto" e "baixo" para cada métrica.

Questão 3: Clusterização: Considerando a rede abaixo:



1. Calcule a clusterização local de cada vértice e a clusterização média.

2. Calcule a clusterização global da rede e compare com a média da clusterização local.

3. Calcule a densidade da rede e compare com as clusterizações calculadas.

Questão 4: Closeness: Lembrando que o closeness de um vértice v é dado pela distância média entre v e todos os outros vértices da rede: $c_v = \frac{1}{n-1} \sum_{u \in V, u \neq v} d(u,v)$, onde $d(u,v)$ é a função que determina a distância entre os vértices u e v. Duas métricas globais populares e muito relacionadas ao closeness são:

Average Pairwise Distance (APD):

$$\frac{1}{\binom{n}{2}} \sum_{u \in V} \sum_{v \in V, v > u} d(u,v) \tag{1}$$

Average Path Length (APL):

$$\frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V} d(u,v) \tag{2}$$

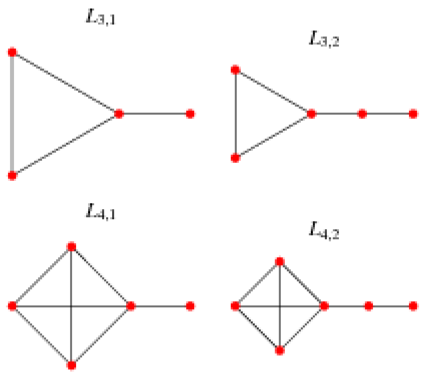
1. Mostre como essas métricas se relacionam. Em particular, reescreva a APD em função de APL. Reescreva uma delas em função do closeness dos vértices.

2. Uma outra propriedade global de redes relacionada ao APD e APL é o diâmetro, $D = \max_{u,v \in V} d(u,v)$. Dê o exemplo mais simples possível de uma rede onde APL e diâmetro são diferentes.

3. Prove ou dê um contra-exemplo para a seguinte afirmação: Para todo grafo G, existe uma constante c independente de $n = |V|$ tal que:

$$\frac{D}{APD} < c \tag{3}$$

4. Uma classe de grafos frequentemente estudada é o grafo do pirulito (*lollipop graph*), $L_{h,k}$, com $h, k > 0$ e inteiros. O grafo pirulito é construído juntando um grafo completo com h vértices com um grafo linha de k vértices (adicionando uma aresta), conforme ilustrado abaixo: Calcule a APD de um grafo $L_{h,k}$ em função de h e k.



Questão 5: Betweenness: Vimos em aula algumas definições para o cálculo do betweenness diante da existência de mais de um caminho mínimo entre dois vértices. Seja $\sigma(i,j)$ e $\sigma_v(i,j)$ o número de caminhos mínimos entre os vértices i e j, e a quantidade destes caminhos que passam por v, respectivamente. Seja $b_v(i,j)$ a contribuição do par (i,j) para o betweenness do vértice v. Em particular, temos as seguintes definições:

1. $b_v(i,j) = 1$ se $\sigma_v(i,j) > 0$. Ou seja, contamos 1 quando v ocorre em ao menos um caminho mínimo entre i e j.

2. $b_v(i,j) = \frac{\sigma_v(i,j)}{\sigma(i,j)}$. Ou seja, a carga de uma unidade trazida pelo par (i,j) é dividida pelo número de caminhos mínimos que conecta os dois vértices.

3. $b_v(i,j) = 1$ se $\sigma_v(i,j) = \sigma(i,j)$. Ou seja, contamos 1 quando v ocorre exatamente em todos os caminhos mínimos entre i e j.

Para cada classe de grafos abaixo, determine analiticamente o betweenness de cada vértice de acordo com cada definição acima. O valor deve ser expresso e função de $n_1 = |V_1|, n_2 = |V_2|, \ldots$:

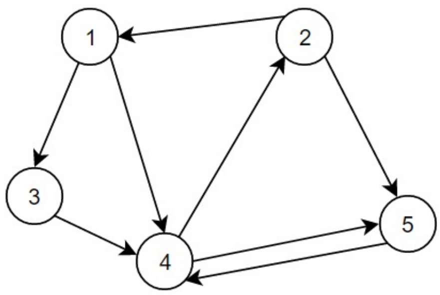
- Grafo completo com $|V| = n$ vértices.
- Grafo bipartido completo, com vértices V_1 e V_2, onde $V = V_1 \cup V_2$.
- Grafo tripartido completo, com vértices V_1, V_2 e V_3, onde $V = V_1 \cup V_2 \cup V_3$.
- Grafo k-partido completo, com vértices $V_1, \ldots, V_k$, onde $V = \bigcup_{i=1}^{k} V_i$.

Questão 6: *PageRank*: Vimos em aula que a centralidade de autovetor pode ser obtida pela fórmula recursiva $x_{t+1} = A^T x_t$, onde A é a matriz de adjacência da rede (grafo direcionado) e A^T sua transposta. Lembrando que se u é um autovetor da matriz A com autovalor λ, então temos que $Au = \lambda u$. Logo, considerando a recursão, no limite quando $t \to \infty$ esperamos que $x^* = A^T x^*$, e logo podemos concluir que x^* é um autovetor da matriz A^T com autovalor $\lambda = 1$.

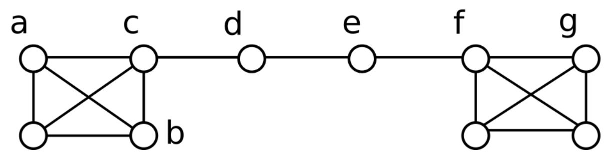
Considere uma versão simplificada do PageRank onde $\alpha = 1, \beta = 0$, com matriz de adjacência A (grafo direcionado), e uma matriz diagonal D com os graus dos vértices, tal que $D_{ii} = d_i$ (o grau do vértice i) e $D_{ij} = 0$, se $i \neq j$. Mostre como utilizar a propriedade acima para calcular a centralidade de PageRank

Questão 7: Passeios aleatórios: Seja $\pi_i(t) = n_i(t)/t$ a fração de vezes que o vértice i é visitado por um passeio aleatório depois de t passos, onde $n_i(t)$ é o número de visitas feita pelo passeio ao vértice i após t passos. Seja $\pi_i = \lim_{t\to\infty} \pi_i(t)$ o limite da fração de visitas, também chamado de estado estacionário do passeio aleatório.

1. Mostre que o estado estacionário de um passeio aleatório em um grafo G não-direcionado é dado por $\pi_i = \frac{d_i}{2m}$, onde d_i é o grau do vértice i e m é o número de arestas do grafo. Sugestão: aplique as equações de fluxo (vista em aula) de forma genérica, isso é, sem determinar numericamente os vizinhos e graus de cada vértice.
2. Calcule o estado estacionário de um passeio aleatório no grafo abaixo utilizando as equações de fluxo. Note que o grafo é direcionado. Caso fosse não-direcionado, qual seria o estado estacionário? Compare os resultados.



Questão 8: Similaridade entre vértices: O grafo *barbell* $B_{h,k}$ é formado por duas cliques de tamanho h ligadas por um caminho com k vértices. Considere o grafo $B_{4,2}$ abaixo para responder os itens seguintes.



1. Calcule as similaridades de Jaccard e do cosseno entre os pares de vértices $\{a, b\}$,$\{b, c\}$,$\{c, d\}$,$\{d, e\}$.

2. O conceito de *identidade estrutural* considera similares vértices cujas vizinhanças tenham padrões de conexões parecidos, mesmo que distantes um do outro no grafo. Considerando as similaridades calculadas acima, as métricas de Jaccard e do cosseno capturam bem a identidade estrutural do grafo? Por quê?

3. Um automorfismo de um grafo $G = (V, E)$ é uma função bijetora $\sigma : V \to V$ que preserva as arestas do grafo. Ou seja, $(u, v) \in E$ se e somente se $(\sigma(u), \sigma(v)) \in E$. Um a automorfismo que mapeia um vértice no outro (ou seja, $\sigma(u) = v$ e $\sigma(v) = u$) é o conceito mais rígido de identidade estrutural, pois significa que estruturalmente os dois vértices são equivalentes. Determine na rede do barbell acima os pares de vértices que podem ter sua identidade trocada e ainda assim fazerem parte de um automorfismo.