

Redes Complexas - CPS765 - 2020.2

Primeira Lista de Exercícios

Pedro Maciel Xavier

116023847

10 de novembro de 2020



Questão 1.: Matriz de Adjacência

Podemos supor, neste caso, que as matrizes em questão vivem em um espaço vetorial construído sobre o semianel $(\mathcal{B}, \vee, \wedge)$ em vez de $(\mathbb{R}, +, \cdot)$, onde $\mathcal{B} = \{0, 1\}$. Assim, as entradas das matrizes serão sempre 0 ou 1 e as operações usuais de soma e multiplicação são substituídas pela disjunção e pela conjunção lógica, respectivamente.

a) A fim de obter uma expressão para a alcançabilidade em k passos do vértice i ao j , dado pela entrada $\mathbf{B}_{i,j}^{(k)}$ vamos empregar um raciocínio indutivo. É claro que a alcançabilidade em 0 passos é dada pela matriz identidade \mathbf{I} , uma vez que só é possível chegar ao vértice em que já encontramos. O caso para um único passo é dado pela matriz de adjacências \mathbf{A} , trivialmente. Logo, $\mathbf{B}^{(0)} = \mathbf{I}$ e $\mathbf{B}^{(1)} = \mathbf{A}$. Vamos supor, por hipótese de indução, que a matriz $\mathbf{B}^{(k)} \in \mathcal{B}^{n \times n}$ representa a alcançabilidade em exatamente k passos, isto é, se existe um caminho de comprimento k ligando o vértice i ao vértice j , então $\mathbf{B}_{i,j}^{(k)} = 1$. Caso contrário, $\mathbf{B}_{i,j}^{(k)} = 0$. Para saber se existe um caminho de tamanho $k + 1$ entre os vértices i e j é preciso que exista um caminho de tamanho k entre i e algum vértice ξ assim como ξ deve ser incidente em j . Portanto,

$$\mathbf{B}_{i,j}^{(k+1)} = \bigvee_{\xi=1}^n \mathbf{B}_{i,\xi}^{(k)} \wedge \mathbf{A}_{\xi,j}$$

de onde concluímos quem, para todo $k \geq 1$, $\mathbf{B}^{(k+1)} = \mathbf{B}^{(k)} \mathbf{A}$. O resultado é dado pelo produto usual de matrizes induzido pelo semianel booleano. Logo, escrevemos $\mathbf{B}^{(k)} = \mathbf{A}^k$.

b) Seguindo raciocínio semelhante, dizemos que i alcança j em k ou menos passos se $\mathbf{B}_{i,j}^{(\xi)} = 1$ para algum $0 \leq \xi \leq k$. Isto é,

$$\mathbf{C}_{i,j}^{(k)} = \mathbf{B}_{i,j}^{(0)} \vee \mathbf{B}_{i,j}^{(1)} \vee \mathbf{B}_{i,j}^{(2)} \dots \vee \mathbf{B}_{i,j}^{(k)} = \bigvee_{\xi=0}^k \mathbf{B}_{i,j}^{(\xi)}$$

resultado que, por conta do espaço onde as matrizes se encontram, é caracterizado pela soma usual. Ou seja, $\mathbf{C}^{(k)} = \sum_{\xi=0}^k \mathbf{B}^{(\xi)}$.

c) Análise da complexidade:

$\mathbf{B}^{(k)}$ - A multiplicação usual de matrizes tem custo $O(n^3)$. Como temos de calcular este produto $k - 1$ vezes, temos uma complexidade assintótica total de ordem $O(n^3 k)$.

$\mathbf{C}^{(k)}$ - A soma de matrizes possui complexidade $O(n^2)$. Contando as $k - 1$ somas temos um total de $O(n^2 k)$ para esta etapa. Se recalculamos $\mathbf{B}^{(k)}$ a cada passo, a complexidade das multiplicações segue uma progressão aritmética em k , totalizando $O(n^3 k^2)$. Se aproveitamos a matriz anterior a cada soma, podemos realizar este processo em tempo $O(n^3 k)$. O termo quadrático em n é de ordem inferior e pode ser omitido em ambos os casos.

d) Seguindo o conselho de multiplicar diferentemente, apresento duas abordagens para reduzir a complexidade do cálculo de $\mathbf{B}^{(k)}$ e $\mathbf{C}^{(k)}$. A primeira, se aplica a um grafo qualquer e se baseia na seguinte relação:

$$\mathbf{A}^k = \begin{cases} \mathbf{I} & \text{para } k = 0 \\ \left(\mathbf{A}^{\frac{k}{2}}\right)^2 & \text{para } k \text{ par} \\ \left(\mathbf{A}^{\frac{k-1}{2}}\right)^2 * \mathbf{A} & \text{para } k \text{ ímpar} \end{cases}$$

para $k \geq 0$. Em geral, esta relação vale para qualquer operação $*$ associativa e, portanto, utilizaremos para o cálculo das potências de matrizes. Isso nos traz complexidade $O(\log k)$ nesta tarefa. Com este aprimoramento, somos capazes de calcular $\mathbf{B}^{(k)}$ em tempo $O(n^3 \log k)$ enquanto $\mathbf{C}^{(k)}$ sai por $O(n^3 \log k!) = O(n^3 k \log k)$. Apesar do ganho no cálculo de $\mathbf{B}^{(k)}$, simplesmente aplicar este método requer calcular cada ξ -ésima potência de \mathbf{A} . É melhor, portanto, multiplicar por \mathbf{A} e somar ao resultado iterativamente, com custo $O(n^3 k + n^2 k) = O(n^3 k)$.

Ainda podemos fazer melhor em alguns casos, dadas algumas condições. Vamos retornar ao espaço euclidiano usual $\mathbb{R}^{n \times n}$, pagando um custo $O(n^2)$ ao final do cálculo para definir cada entrada de $\mathbf{B}^{(k)}$ e $\mathbf{C}^{(k)}$ como 0 ou 1 verificando se cada elemento da matriz é ou não nulo, respectivamente. Se $\det(\mathbf{I} - \mathbf{A}) = (-1)^n \det(\mathbf{A} - \mathbf{I}) \neq 0$, isto é, para cada autovalor λ de \mathbf{A} temos que $\lambda \neq 1$, podemos dizer que

$$\sum_{\xi=0}^k \mathbf{A}^\xi = (\mathbf{I} - \mathbf{A})^{-1}(\mathbf{I} - \mathbf{A}^{k+1})$$

A inversão da matriz $\mathbf{I} - \mathbf{A}$ pode ser feita em tempo $O(n^3)$ pela eliminação de *Gauss-Jordan*. Isso nos permite calcular $\mathbf{C}^{(k)}$ em tempo $O(n^3) + O(n^3 \log k)$, ou seja, $O(n^3 \log k)$.

Temos ainda um caso ainda mais específico, para grafos não-direcionados. Neste caso, a estrutura confere $\mathbf{A} = \mathbf{A}^T$. O Teorema Espectral nos garante, portanto, que a matriz \mathbf{A} é diagonalizável e, além disso, a simetria permite encontrar a forma $\mathbf{A} = \mathbf{S}\mathbf{\Lambda}\mathbf{S}^{-1}$ em tempo $O(n^3)$ através da transformação de *Householder* seguida da aplicação do algoritmo *QR*. Em seguida, calculamos $\mathbf{B}^{(k)} \sim \mathbf{A}^k = \mathbf{S}\mathbf{\Lambda}^k\mathbf{S}^{-1}$ em tempo $O(n^3 + n \log k)$, uma vez que basta calcular a potência de cada uma das n entradas da diagonal principal de $\mathbf{\Lambda}$, o que possui complexidade $O(\log k)$ segundo o método visto acima. Por fim, isso também se aplica ao cálculo de $\mathbf{C}^{(k)}$, que pelo método da soma geométrica de matrizes descrito acima pode ser feito em tempo $O(n^3 + n \log k)$. Isso se verifica também por outra propriedade da forma diagonal de \mathbf{A} , uma vez que

$$\sum_{\xi=0}^k \mathbf{A}^\xi = \mathbf{S} \left[\sum_{\xi=0}^k \mathbf{\Lambda}^\xi \right] \mathbf{S}^{-1} = \mathbf{S} \begin{bmatrix} \sum_{\xi=0}^k \lambda_1^\xi & & \\ & \ddots & \\ & & \sum_{\xi=0}^k \lambda_n^\xi \end{bmatrix} \mathbf{S}^{-1} = \mathbf{S} \begin{bmatrix} \frac{1 - \lambda_1^{k+1}}{1 - \lambda_1} & & \\ & \ddots & \\ & & \frac{1 - \lambda_n^{k+1}}{1 - \lambda_n} \end{bmatrix} \mathbf{S}^{-1}$$

onde fica clara a condição de que $\lambda \neq 1$.

Questão 2.: Grau médio e densidade

Para analisar o comportamento das duas propriedades (grau médio \bar{g} e densidade ρ) vamos escrever cada uma em função da outra. Partimos das expressões

$$\bar{g} = \frac{2m}{n} \text{ e } \rho = \frac{2m}{n(n-1)}$$

de onde definimos as funções

$$\bar{g}(\rho) = \rho(n-1) \text{ e } \rho(\bar{g}) = \frac{\bar{g}}{(n-1)}$$

O grau médio é, portanto, uma dilatação da densidade por um fator $(n-1)$ para $n \in \mathbb{N}$ qualquer. Possui comportamento monotônico crescente como função de ρ . Desta forma, para um grafo qualquer de n vértices, independentemente de sua configuração de arestas, sabemos que o grau médio e a densidade estão relacionados de maneira linear. Mais do que isso, dizemos que $\bar{g} \propto \rho$. Isso garante que altas densidades levam a um alto grau médio da rede, e vice-versa.

Questão 3.: clusterização

1 Calculando a clusterização local de cada vértice do grafo:

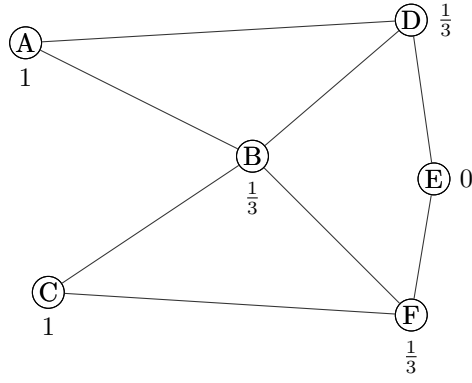


Figura 1: Clusterização local

A clusterização média será, portanto,

$$\frac{1}{6} \sum_{i=A}^F c_i = \frac{1}{2} = 0.5$$

2 Contando as triplas e os triângulos:

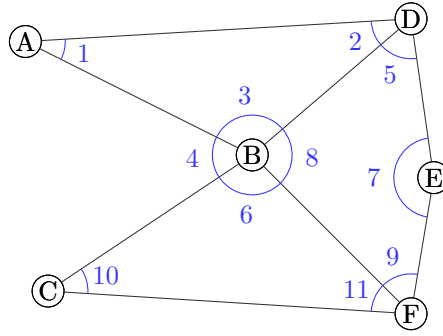


Figura 2: Triplas numeradas

Temos, então, 11 triplas e 2 triângulos ($\triangle ABD$ e $\triangle BCF$). Logo, a clusterização global é dada por

$$\frac{3 \times \text{n}^\circ \text{ de triângulos}}{\text{n}^\circ \text{ de triplas}} = \frac{6}{11} = 0.\overline{54}$$

3 A densidade da rede é dada por simplesmente $\rho = \frac{2m}{n(n-1)}$, ou seja, $\rho = \frac{16}{30} = 0.5\overline{3}$. Observa-se que o valor da densidade se encontra entre o da clusterização média e o da clusterização global, estando mais próximo desta última.

Questão 4.: Closeness

Lembrando que o *closeness* de um vértice v é dado pela expressão

$$c_v = \frac{1}{n-1} \sum_{u \neq v \in V} d(u, v)$$

onde $d(u, v)$ é a distância entre u e v definimos as métricas globais:

- *Average Pairwise Distance* (APD):

$$\text{APD}(V, E) = \frac{1}{\binom{n}{2}} \sum_{u \in V} \sum_{v > u \in V} d(u, v)$$

- *Average Path Lenght* (APL:)

$$\text{APL}(V, E) = \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V} d(u, v)$$

É importante notar que a métrica APD só se aplica a grafos não-direcionados, uma vez que os pares (u, v) do somatório sempre tem $v > u$. Aplicar sobre grafos direcionados geraria assimetria conforme a numeração dos vértices.

1 Reescrevendo APD:

$$\begin{aligned} \text{APD}(V, E) &= \frac{1}{\binom{n}{2}} \sum_{u \in V} \sum_{v > u \in V} d(u, v) \\ &= \frac{2}{n(n-1)} \sum_{u \in V} \sum_{v > u \in V} d(u, v) \\ &= \frac{1}{n(n-1)} \sum_{u \in V} \left[\sum_{v > u \in V} d(u, v) + \sum_{v < u \in V} d(u, v) \right] \\ &= \frac{1}{n(n-1)} \sum_{u \in V} \left[\sum_{v > u \in V} d(u, v) + \sum_{v < u \in V} d(u, v) + \sum_{v = u \in V} d(u, v) \right] \\ &= \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \in V} d(u, v) \\ &= \text{APL}(V, E) \end{aligned}$$

Aqui assumimos que $d(u, v) = d(v, u) \forall u, v \in V$ e $d(u, u) = 0 \forall u \in V$. Esta última suposição apenas explicita que *loops* não são levados em consideração, deixando assim a definição compatível com o termo $n(n-1)$ que divide o total.

Vamo agora reescrever a métrica APD em função do *closeness* dos vértices:

$$\begin{aligned} \text{APD}(V, E) &= \frac{1}{n(n-1)} \sum_{u \in V} \left[\sum_{v > u \in V} d(u, v) + \sum_{v < u \in V} d(u, v) \right] \\ &= \frac{1}{n(n-1)} \sum_{u \in V} \sum_{v \neq u \in V} d(u, v) \\ &= \frac{1}{n} \sum_{u \in V} c_u \end{aligned}$$

onde c_u é o *closeness* do vértice u . Vale notar que isso diz que APD e APL podem ser entendidos como o *closeness* médio.

2 Se o APL e o diâmetro de uma rede são iguais, então

$$\sum_{u \in V} \sum_{v \in V} d(u, v) = n(n-1) \max_{u, v \in V} d(u, v) \iff d(i, j) = \max_{u, v \in V} d(u, v) \quad \forall i, j \in V, i \neq j$$

para $n \leq 2$, a relação vale. Escolhendo $n = 3$, portanto, encontramos uma configuração onde a equação não é satisfeita

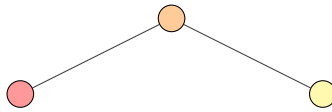


Figura 3: O menor contra-exemplo

já que temos distâncias entre pares de valores distintos, sendo este o menor grafo conexo não-completo de n vértices. Vale lembrar que em todo momento consideramos $d(i, j) = d(j, i)$ assim como $d(i, i) = 0$.

3 Sabemos que o valor da razão entre o diâmetro e a métrica APD de um grafo qualquer de n vértices é limitada por

$$1 \leq \frac{D(V, E)}{APL(V, E)} \leq \binom{n}{2}$$

Isto segue diretamente da equivalência das normas $\|\mathbf{x}\|_\infty \leq \|\mathbf{x}\|_1 \leq \xi \|\mathbf{x}\|_\infty$ com $\mathbf{x} \in \mathbb{R}^\xi$. De fato, para grafos completos e linhas a razão é limitada por constantes.

No entanto, isto não vale para qualquer tipo de grafo. Seja $\mathcal{L}_{p,q}$ o grafo de *Barbell* cujas duas cliques \mathcal{K}_p e são ligadas por uma linha de q vértices. O diâmetro desta rede é $q + 2$, enquanto o valor do APD é ?

4 Seja $\mathcal{L}_{h,k}$ um grafo pirulito fruto da junção da clique \mathcal{K}_h com um caminho simples \mathcal{C}_k onde $n = k + h$ é o total de vértices.

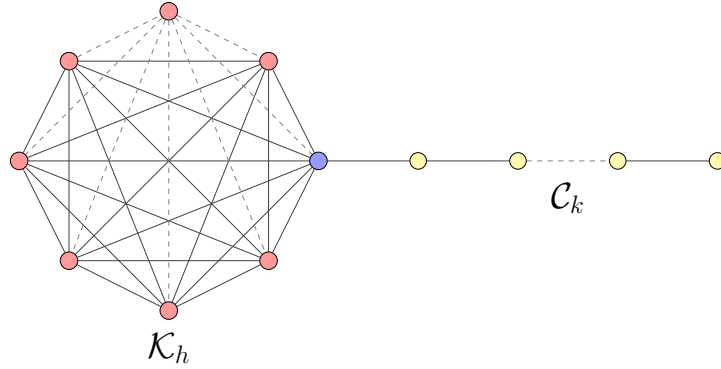


Figura 4: Um Grafo Pirulito

Lembrando que

$$APD(V, E) = \frac{1}{\binom{n}{2}} \sum_{u \in V} \sum_{v > u \in V} d(u, v) = \frac{1}{n-1} \sum_{u \in V} c_u$$

onde c_u é o *closeness* do vértice u .

Vamos primeiro dividir o grafo em duas componentes conexas. Seja $\hat{\mathcal{K}}$ o conjunto (em vermelho) dos vértices da clique, exceto o vértice da ponte (em azul). Seja também $\hat{\mathcal{C}}$ a junção do grafo linha (em amarelo) com o vértice na ponte. Vamos numerar os vértices em $\hat{\mathcal{C}}$ de 1 até $k + 1$, da esquerda para a direita.

Agora, vamos calcular o *closeness* de um vértice κ qualquer em $\hat{\mathcal{K}}$.

$$\begin{aligned} c_\kappa &= \frac{1}{h-2} \sum_{i=1}^{h-2} 1 + \frac{1}{k+1} \sum_{i=1}^{k+1} i \\ &= 1 + \frac{(k+1)(k+2)}{2(k+1)} = \frac{k+4}{2} \end{aligned}$$

Em seguida, o *closeness* do ξ -ésimo vértice de $\hat{\mathcal{C}}$, da maneira que os numeramos.

$$c_\xi$$

Questão 5.: *Betweenness*

Seja $b_v(i, j)$ a contribuição do par (i, j) para o *betweenness* do vértice v . Temos as seguintes definições:

1. $\hat{b}_v(i, j) = \mathbb{I}\{\sigma_v(i, j) > 0\}$
2. $b_v(i, j) = \frac{\sigma_v(i, j)}{\sigma(i, j)}$
3. $\check{b}_v(i, j) = \mathbb{I}\{\sigma_v(i, j) = \sigma(i, j)\}$

Podemos pensar as definições 1 e 3 como sendo as versões otimista e pessimista da definição 2, respectivamente.

1 Em um grafo completo com $|V| = n$ vértices, sempre existe uma aresta entre um par (i, j) . Portanto, todo caminho mínimo entre i e j é único e possui comprimento 1. Desta forma, $b_v(i, j) = 0$ independentemente da métrica utilizada. O *betweenness* de cada vértice é, portanto, igual a zero.

2 Em um grafo bipartido completo, com conjuntos disjuntos de vértices V_1 e V_2 , onde $V = V_1 \cup V_2$, temos dois casos a serem analisados separadamente. Para um dado par de vértices (i, j) temos que, se $v \in V_\xi$ e $\bar{V}_\xi = V - V_\xi$ com $\xi = 1, 2$ então

$$b_{v \in V_\xi}(i, j) = \begin{cases} \frac{1}{|\bar{V}_\xi|} & \text{se } i, j \in \bar{V}_\xi \\ 0 & \text{caso contrário} \end{cases}$$

uma vez que

O *betweenness* de um vértice v é, portanto, segundo cada métrica normalizada,

$$\begin{aligned} \langle \mathbb{b}(v \in V_\xi) \rangle &= \frac{1}{\binom{n}{2}} \sum_{i, j \in V} b_{v \in V_\xi}(i, j) \\ &= \frac{1}{\binom{n}{2}} \left[\sum_{i, j \in V_\xi} b_{v \in V_\xi}(i, j) + \sum_{\substack{i \in V_\xi \\ j \in \bar{V}_\xi}} b_{v \in V_\xi}(i, j) + \sum_{\substack{i \in \bar{V}_\xi \\ j \in V_\xi}} b_{v \in V_\xi}(i, j) + \sum_{i, j \in \bar{V}_\xi} b_{v \in V_\xi}(i, j) \right] \\ &= \frac{1}{\binom{n}{2}} \sum_{i, j \in \bar{V}_\xi} \frac{1}{|\bar{V}_\xi|} = \frac{\bar{n}_\xi(\bar{n}_\xi - 1)}{n_\xi n(n - 1)} \end{aligned}$$

$$\therefore \langle \hat{\mathbb{b}}(v \in V_\xi) \rangle = \frac{1}{\binom{n}{2}} \sum_{i, j \in V} \hat{b}_{v \in V_\xi}(i, j) = \frac{1}{\binom{n}{2}} \sum_{i, j \in \bar{V}_\xi} 1 = \frac{\bar{n}_\xi(\bar{n}_\xi - 1)}{n(n - 1)}$$

$$\therefore \langle \check{\mathbb{b}}(v \in V_\xi) \rangle = \frac{1}{\binom{n}{2}} \sum_{i, j \in V} \check{b}_{v \in V_\xi}(i, j) = \frac{1}{\binom{n}{2}} \sum_{i, j \in \bar{V}_\xi} 0 = 0$$

onde $n_\xi = |V_\xi|$ e $\bar{n}_\xi = |\bar{V}_\xi|$.

3 Para o grafo tripartido completo com $V = V_1 \cup V_2 \cup V_3$ seguiremos uma linha de pensamento semelhante ao caso anterior, do grafo bipartido. No entanto, vamos tomar uma análise um tanto mais ambiciosa. Seja $G(V, E)$ um grafo k -partido completo, ou seja, $V = V_1 \cup V_2 \cup \dots \cup V_k$. A contribuição $b_v(i, j)$ de um par $(i, j) \in V$ para o *betweenness* do vértice $v \in V$ só existe quando $i, j \in V_\xi$ e $v \notin V_\xi$ para algum $1 \leq \xi \leq k$, onde $\bigcup_{\xi=1}^k V_\xi = V$ e $V_\xi \cap V_\zeta = \emptyset$ se $\xi \neq \zeta$. Uma vez atendida esta condição, temos que um dos caminhos mínimos entre i e j passará por v , portanto $\sigma_v(i, j) = 1$. Caso contrário, $\sigma_v(i, j) = 0$. De qualquer forma, temos $|V_\zeta|$ caminhos mínimos entre

$i, j \in V_\xi$ para cada $V_\zeta, \zeta \neq \xi$. Portanto, $\sigma(i, j) = \sum_{\zeta \neq \xi} |V_\zeta|$ quando $i, j \in V_\xi$. Seja $\overline{V}_\xi = V - V_\xi$, $\xi = 1, 2, \dots, k$. Logo, para um v3rtice v em algum $V_\zeta \subseteq \overline{V}_\xi$

$$b_{v \in \overline{V}_\xi}(i, j) = \begin{cases} \frac{1}{|\overline{V}_\xi|} & \text{se } i, j \in V_\xi \\ 0 & \text{caso contr3rio} \end{cases}$$

Portanto, as m3tricas normalizadas para um v3rtice $v \in V_\zeta$ em um grafo completo k -partido s3o expressas por

$$\begin{aligned} \langle \mathbb{b}(v \in V_\zeta) \rangle &= \frac{1}{\binom{n}{2}} \sum_{i, j \in V} b_{v \in \overline{V}_\xi}(i, j) \\ &= \frac{1}{\binom{n}{2}} \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^k \sum_{i, j \in V_\xi} b_{v \in \overline{V}_\xi}(i, j) \\ &= \frac{1}{\binom{n}{2}} \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^k \sum_{i, j \in V_\xi} \frac{1}{|\overline{V}_\xi|} \\ &= \frac{1}{\binom{n}{2}} \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^k \sum_{i, j \in V_\xi} \frac{1}{n - n_\xi} \\ &= \frac{1}{\binom{n}{2}} \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^k \frac{1}{n - n_\xi} \cdot \frac{n_\xi(n_\xi - 1)}{2} \\ &= \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^k \frac{1}{n - n_\xi} \cdot \frac{n_\xi(n_\xi - 1)}{n(n - 1)} \end{aligned}$$

o que, de fato, 3 compat3vel com o caso em que $k = 2$ (grafo bipartido completo) trocando-se V_ξ por \overline{V}_ξ (e, conseq3entemente, n_ξ por \overline{n}_ξ). Vale lembrar que $\overline{n}_\xi = n - n_\xi$.

Por fim, aplicando a f3rmula derivada acima para $k = 3$ (grafo tripartido completo) obtemos

$$\begin{aligned} \langle \mathbb{b}(v \in V_\zeta) \rangle &= \frac{1}{\binom{n}{2}} \sum_{i, j \in V} b_{v \in \overline{V}_\xi}(i, j) \\ &= \sum_{\substack{\xi=1 \\ \xi \neq \zeta}}^3 \frac{1}{n - n_\xi} \cdot \frac{n_\xi(n_\xi - 1)}{n(n - 1)} \\ &= \frac{1}{n(n - 1)} \cdot \left[\frac{n_\phi(n_\phi - 1)}{n - n_\phi} + \frac{n_\psi(n_\psi - 1)}{n - n_\psi} \right] \\ \therefore \langle \hat{\mathbb{b}}(v \in V_\zeta) \rangle &= \frac{1}{\binom{n}{2}} \sum_{i, j \in V} \hat{b}_{v \in V_\zeta}(i, j) \\ &= \frac{n_\phi(n_\phi - 1) + n_\psi(n_\psi - 1)}{n(n - 1)} \\ \therefore \langle \check{\mathbb{b}}(v \in V_\zeta) \rangle &= \frac{1}{\binom{n}{2}} \sum_{i, j \in V} \check{b}_{v \in V_\zeta}(i, j) = 0 \end{aligned}$$

com $1 \leq \zeta \neq \phi \neq \psi \leq 3$.

Questão 6.: PageRank

1 A centralidade de *PageRank* é a solução para o sistema de equações abaixo:

$$\mathbf{x}_i = \alpha \sum_{j=1}^n \mathbf{A}_{j,i} \frac{\mathbf{x}_j}{\mathbf{d}_j^s} + \frac{1-\alpha}{n}$$

onde \mathbf{d}_j^s é o grau de saída do vértice j . Este termo pode ser absorvido pela matriz de adjacências, dividindo cada j -ésima linha da matriz por \mathbf{d}_j^s , obtendo assim a matriz \mathbf{D} . Além disso, o somatório está representando um produto interno, entre a i -ésima linha de \mathbf{D}^T e o vetor \mathbf{x} , que mede a centralidade de cada vértice. Tomando $\beta = \frac{1-\alpha}{n}$, podemos reescrever a equação como:

$$\mathbf{x} = [\alpha \mathbf{D}^T + \beta \mathbf{1}] \mathbf{x}$$

onde $\mathbf{1} \in \mathbb{R}^{n \times n}$ tem entradas $\mathbb{1}_{i,j} = 1$. Esta equação vale quando levamos em consideração que $\|\mathbf{x}\|_1 = 1$ e que $0 \leq \mathbf{x}_i \leq 1$. Seja $\mathbf{M} = \alpha \mathbf{D}^T + \beta \mathbf{1}$, temos que a solução do sistema \mathbf{x} é o autovetor principal de \mathbf{M} com autovalor correspondente $\lambda = 1$. Como a primeira equação é equivalente a esta última, e cada iteração do método da potência é dada por $\mathbf{x}^{(k+1)} = \mathbf{M}\mathbf{x}^{(k)}$ a menos de uma normalização, podemos calcular o *PageRank* iterativamente, pela relação

$$\mathbf{x}_i^{(k+1)} = \alpha \sum_{j=1}^n \mathbf{A}_{j,i} \frac{\mathbf{x}_j^{(k)}}{\mathbf{d}_j^s} + \frac{1-\alpha}{n}$$

Para tratar todos os nós com igual importância inicial, atribui-se $\mathbf{x}_i^{(0)} = \frac{1}{n}$.

2 Vamos agora calcular o *PageRank* da rede abaixo:

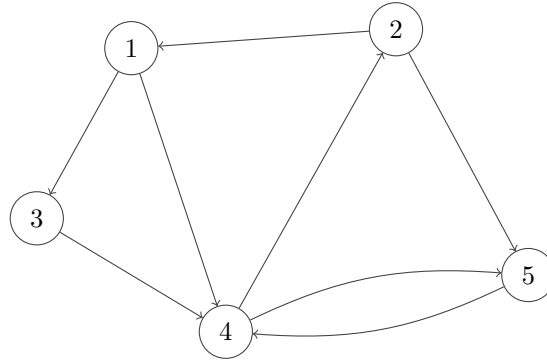


Figura 5: Rede abaixo

Nas tabelas a seguir, temos o valor da centralidade de cada vértice após 5, 10, e 30 iterações, para os valores de α 0.1 e 0.9, respectivamente.

v	5	10	30
1	.189	.189	.189
2	.191	.191	.191
3	.189	.189	.189
4	.228	.228	.228
5	.200	.200	.200

Figura 6: $\alpha = 0.1$

v	5	10	30
1	.103	.104	.104
2	.188	.186	.186
3	.064	.066	.066
4	.371	.371	.370
5	.271	.271	.271

Figura 7: $\alpha = 0.9$

Esta análise permite ver que o algoritmo converge muito rapidamente. Este fato se mostra ainda mais contundente quando se observa que o mais relevante para o resultado é a ordenação hierárquica e não os valores propriamente ditos.

Mesmo assim, acredito que para entender melhor o comportamento do algoritmo ao longo do processo é interessante apresentar os gráficos com a evolução temporal para cada valor de α , na figura a seguir.

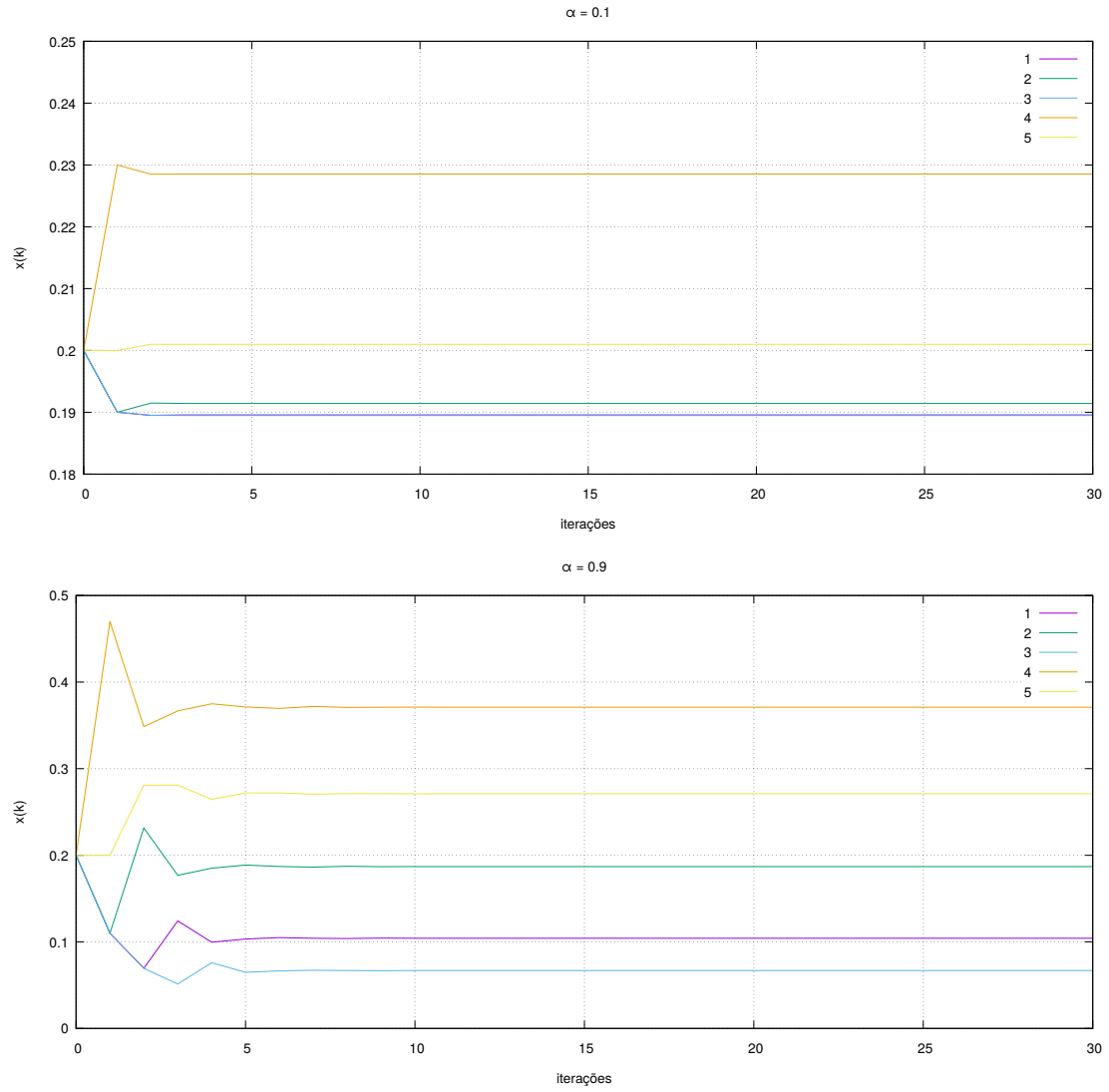


Figura 8: Evolução do *PageRank*

Questão 7.: Similaridade entre vértices

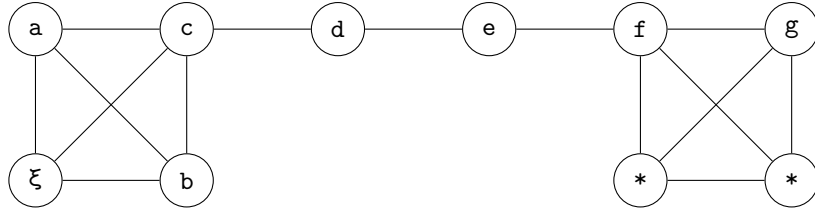


Figura 9: Grafo de *Barbell* $\mathcal{B}_{4,2}$

1 Recapitulando as definições da similaridade de *Jaccard* $\mathcal{J}(i, j)$ e da similaridade do Cosseno $\mathcal{C}(i, j)$ dizemos que

$$\mathcal{J}(i, j) = \frac{|\text{viz } i \cap \text{viz } j|}{|\text{viz } i \cup \text{viz } j|} \text{ e } \mathcal{C}(i, j) = \frac{\langle \mathbf{A}_i, \mathbf{A}_j \rangle}{\|\mathbf{A}_i\| \cdot \|\mathbf{A}_j\|} = \frac{|\text{viz } i \cap \text{viz } j|}{\sqrt{|\text{viz } i| \cdot |\text{viz } j|}}$$

onde $\text{viz } \xi$ denota o conjunto dos vértices na vizinhança de ξ e \mathbf{A}_ξ é a linha correspondente na matriz de adjacências.

Organizando os resultados em uma tabela:

i, j	$\text{viz } i$	$\text{viz } j$	$\text{viz } i \cap \text{viz } j$	$\text{viz } i \cup \text{viz } j$	$\mathcal{J}(i, j)$	$\mathcal{C}(i, j)$
a, b	$\{b, c, \xi\}$	$\{a, c, \xi\}$	$\{c, \xi\}$	$\{a, b, c, \xi\}$.50	.66
b, c	$\{a, c, \xi\}$	$\{a, b, d, \xi\}$	$\{a, \xi\}$	$\{a, b, c, d, \xi\}$.40	.58
c, d	$\{a, b, d, \xi\}$	$\{c, e\}$	\emptyset	$\{a, b, c, d, e, \xi\}$	0	0
d, e	$\{c, e\}$	$\{d, f\}$	\emptyset	$\{c, d, e, f\}$	0	0

2 Não, muito pelo contrário. Como ambas possuem a interseção das vizinhanças no numerador, vértices cujas conexões são idênticas a menos de isomorfismo não serão contemplados por estas duas métricas. Do ponto de vista da *identidade estrutural*, poderíamos afirmar que os vértices d e e são estruturalmente idênticos. Se alguém removesse as etiquetas da Figura 9 e chacoalhasse o grafo um pouco, já seria impossível afirmar com certeza quais vértices estavam marcados por d e e . No entanto, em ambas as métricas este par de vértices foi tratado como plenamente dissemelhante.

3 O grafo de *Barbell* foi colorido na figura abaixo de modo que dois vértices possam ter suas identidades trocadas e ainda pertencer a um automorfismo σ caso tenham sido coloridos com a mesma cor.

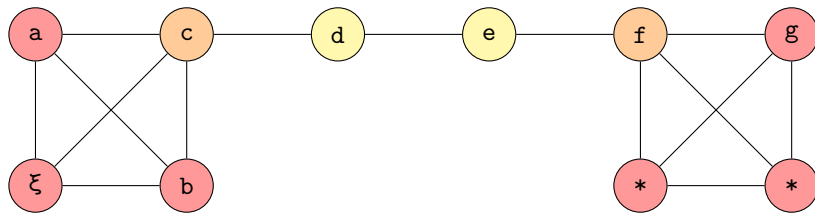


Figura 10: Um mapa para o Automorfismo