Algoritmos para Estimação de Modelos Gráficos

Rodrigo Ribeiro Santos de Carvalho
Orientadora: Florencia Graciela Leonardi

Instituto de Matemática e Estatística
Universidade de São Paulo

Propriedade Local de Markov: Para $G = (V, E)$, existir $ne(v) \subseteq V \setminus v$ tal que para todo $W \subseteq V \setminus ne(v) \cup \{v\}$

$$\mathbb{P}(X_v = a_v | X_{ne(v)} = a_{ne(v)}) = \mathbb{P}(X_v = a_v | X_{ne(v)} = a_{ne(v)}, X_W = a_W)$$

Considerando uma amostra de tamanho n, estimamos $ne(v)$, com $c > 0$, por

$$\widehat{ne}(v) = \underset{W \subseteq V \setminus \{v\}}{\arg\max} \left\{ \sum_{a \in A} \sum_{a_W \in A^W} N(a_v, a_W) \log \widehat{p}(a_v | a_W) - c|A|^{|W|} \log n \right\}$$

Teorema. Para qualquer $c > 0$, $\widehat{ne}(v) \xrightarrow{q.c} ne(v)$, quando $n \to \infty$.

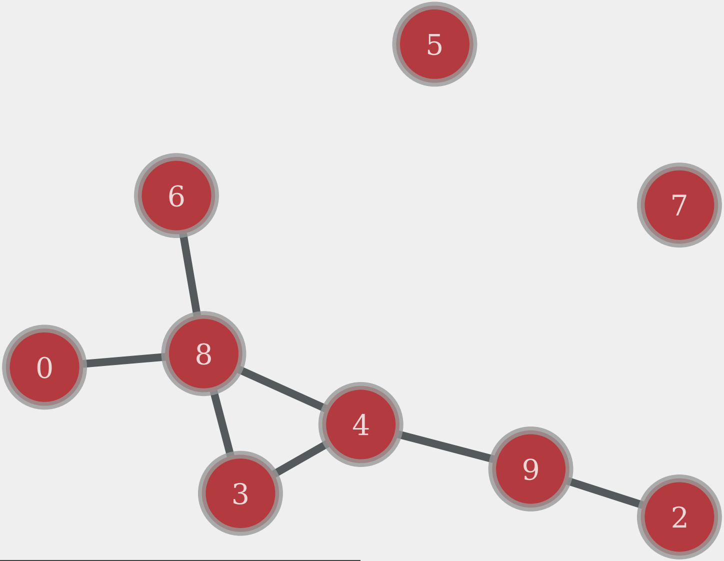
Estima um campo aleatórios de Markov de tipo árvore.

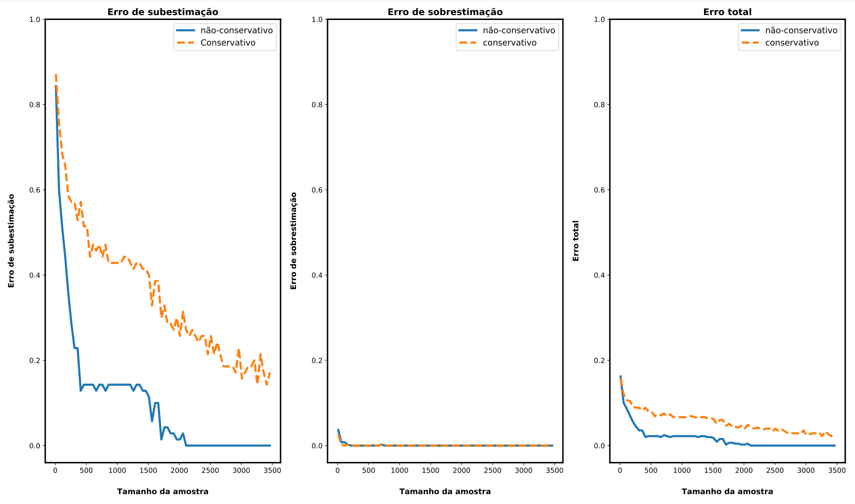
Obtemos a *medida de informação mútua* entre cada par de vértices:

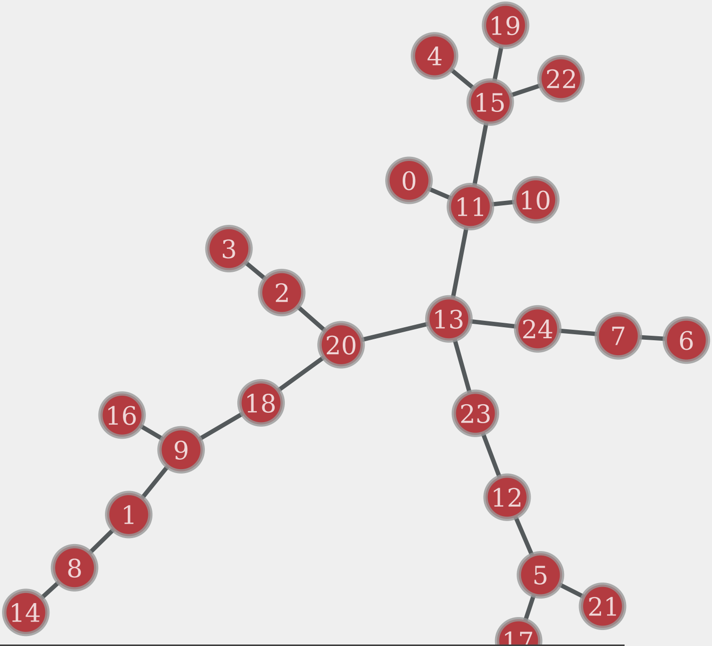
$$\widehat{I}(X_v, X_w) = \sum_{(a_v, a_w) \in A^2} \frac{N(a_v, a_w)}{n} [\log \frac{N(a_v, a_w)}{N(a_v)(N_w)} + \log n]$$

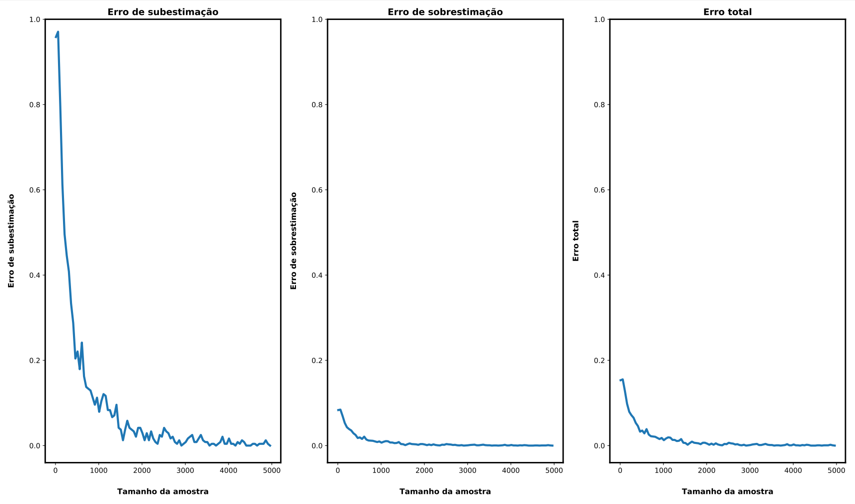
Aplicamos um algoritmo de árvore geradora máxima.

Simulações




Erro de subestimação
Erro de subestimação
Tamanho da amostra
não-conservativo
Conservativo
Erro de sobrestimação
Erro de sobrestimação
Tamanho da amostra
não-conservativo
conservativo
Erro total
Erro total
Tamanho da amostra
não-conservativo
conservativo
0.0
0.2
0.4
0.6
0.8
1.0
0
500
1000
1500
2000
2500
3000
3500


19
4
22
15
0
11
10
3
2
13
24
7
6
20
18
16
9
23
12
1
8
5
21
14
17


Erro de subestimação
Erro de sobrestimação
Erro total
Erro de subestimação
Erro de sobrestimação
Erro total
1.0
0.8
0.6
0.4
0.2
0.0
0
1000
2000
3000
4000
5000
Tamanho da amostra
Tamanho da amostra
Tamanho da amostra

Aplicação