

ANÁLISE DE ROBUSTEZ DO MÉTODO DE INTEGRAÇÃO DE DADOS NERI

Aluno: João Carlos Pandolfi Santana

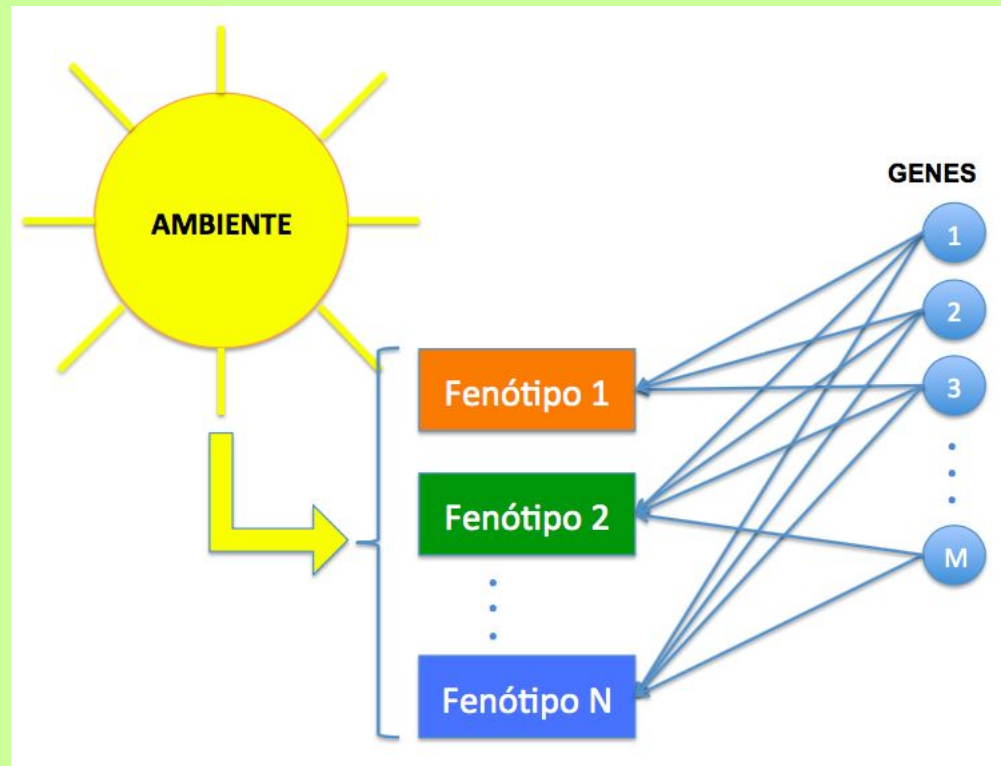
Orientador: Prof. Dr. Sérgio Nery Simões

Serra – ES, Julho de 2017

1. Visão geral
2. Fundamentos conceituais
3. Metodologia proposta
4. Desenvolvimento e Resultados
5. Conclusão
6. Trabalhos futuros

Introdução

- Doenças complexas são poligênicas e multifatoriais, ou seja, além de serem causadas por mutação de mais de um gene, também são afetadas por fatores externos.



Motivação

- Dificuldade do estudo de doenças complexas.
- Método NERI apresenta bons resultados de replicabilidade, mas falta avaliar sua robustez em relação aos genes sementes.

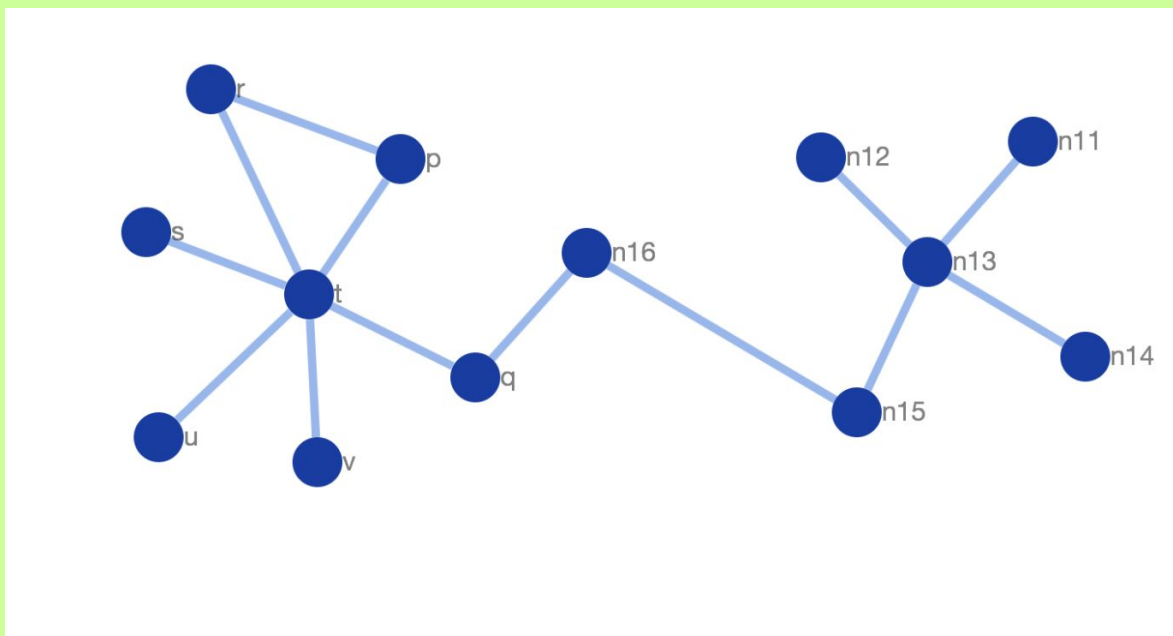
1. Visão Geral - Problema

O Quão dependente é o método NERI dos genes sementes?

Objetivo Geral

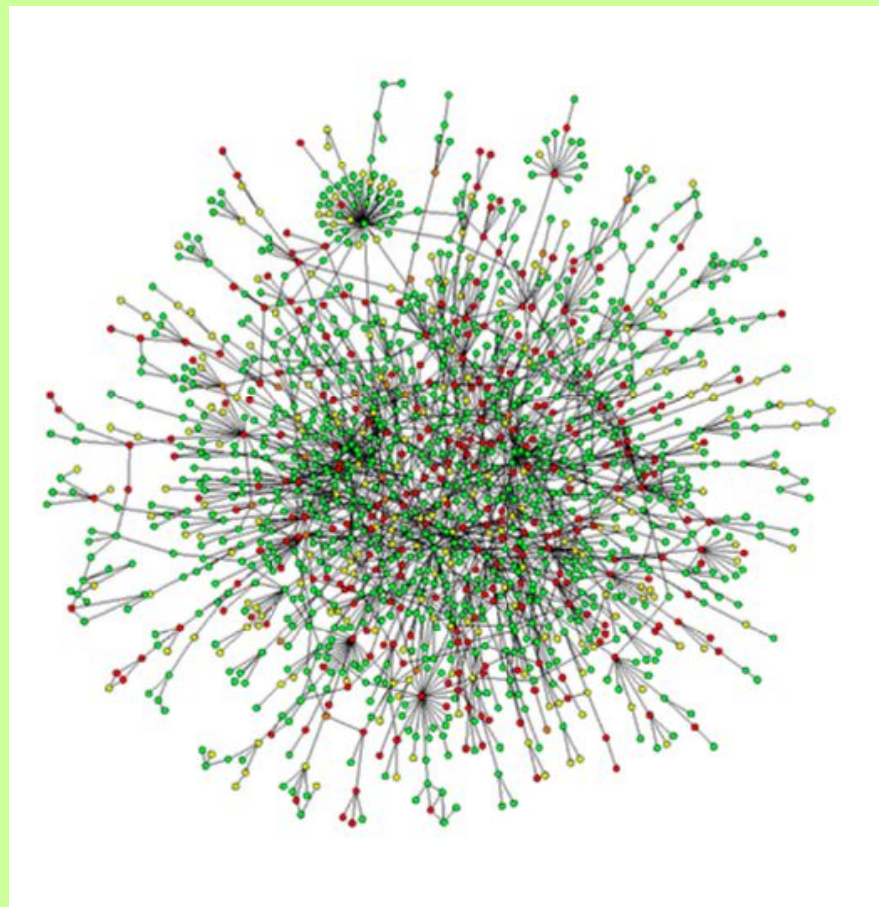
- Analisar robustez do método NERI avaliando o impacto da retirada de alguns genes sementes.

2. Fundamentos conceituais – Grafo e medidas de centralidade



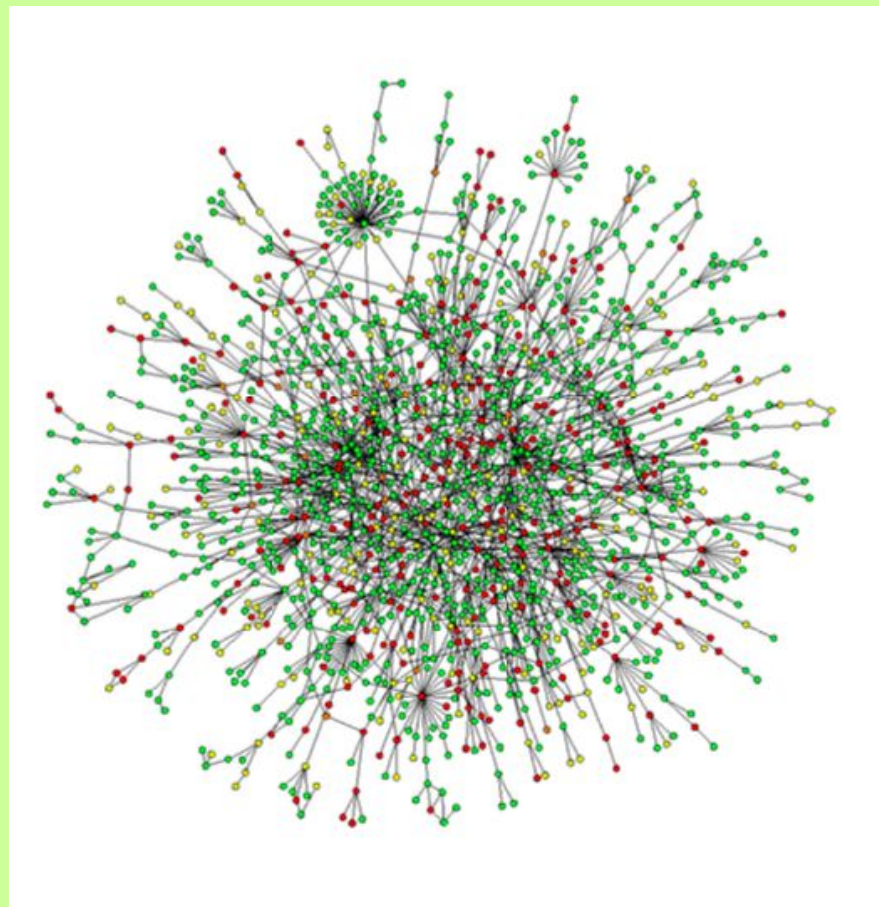
2. Fundamentos conceituais – Redes complexas

- Surge com o estudo de problemas do mundo real.
- Não são modelos simples (ex: regulares nem aleatórios).
- Podem ser modeladas em grafos.

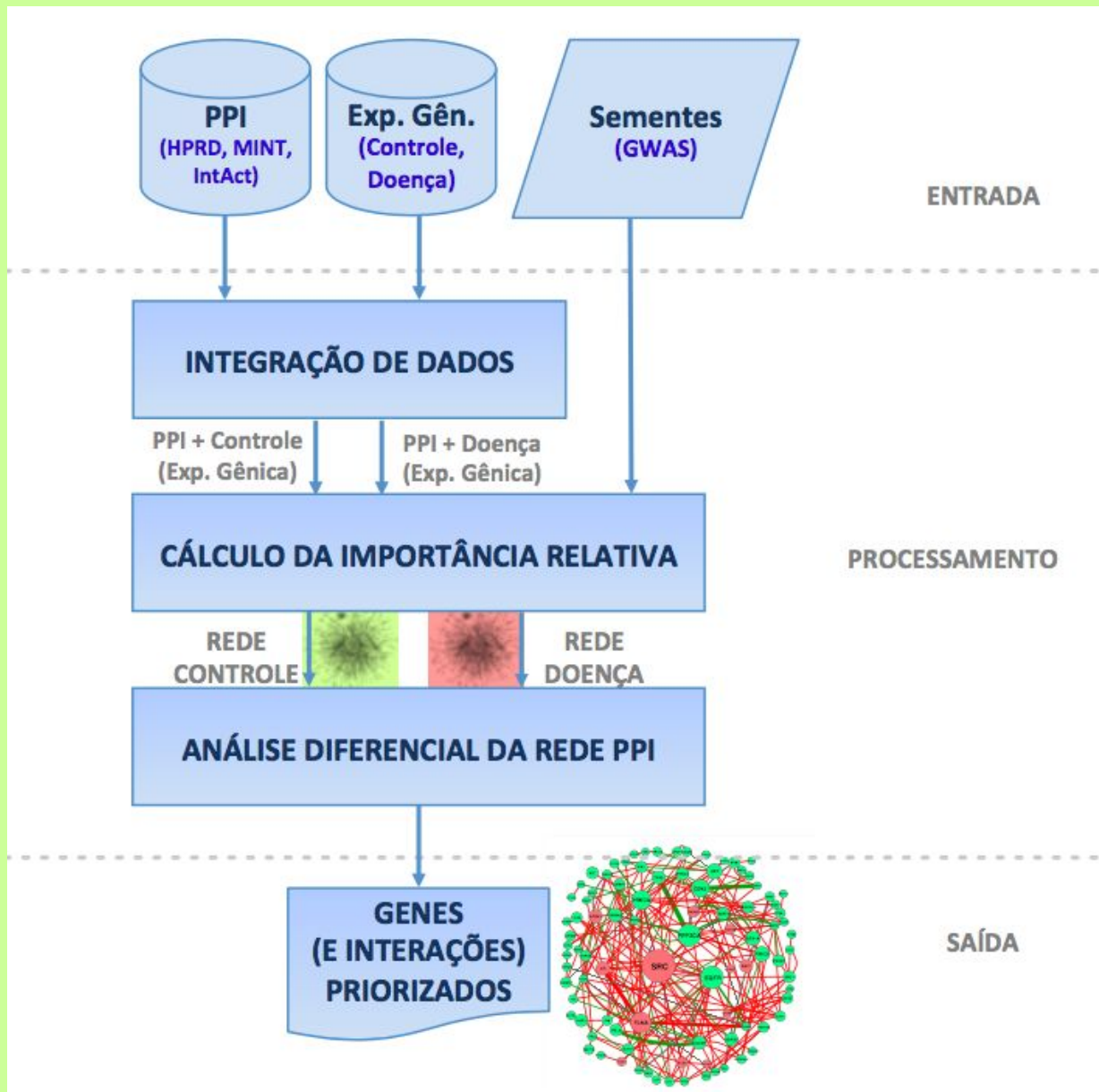


2. Fundamentos conceituais – Redes PPI

- Fornecem informações biológicas sobre potenciais interações entre proteínas.
- São estáticas, atemporais, genéricas e não fornecem informações de causalidade.
- Podem ser modeladas em grafos.

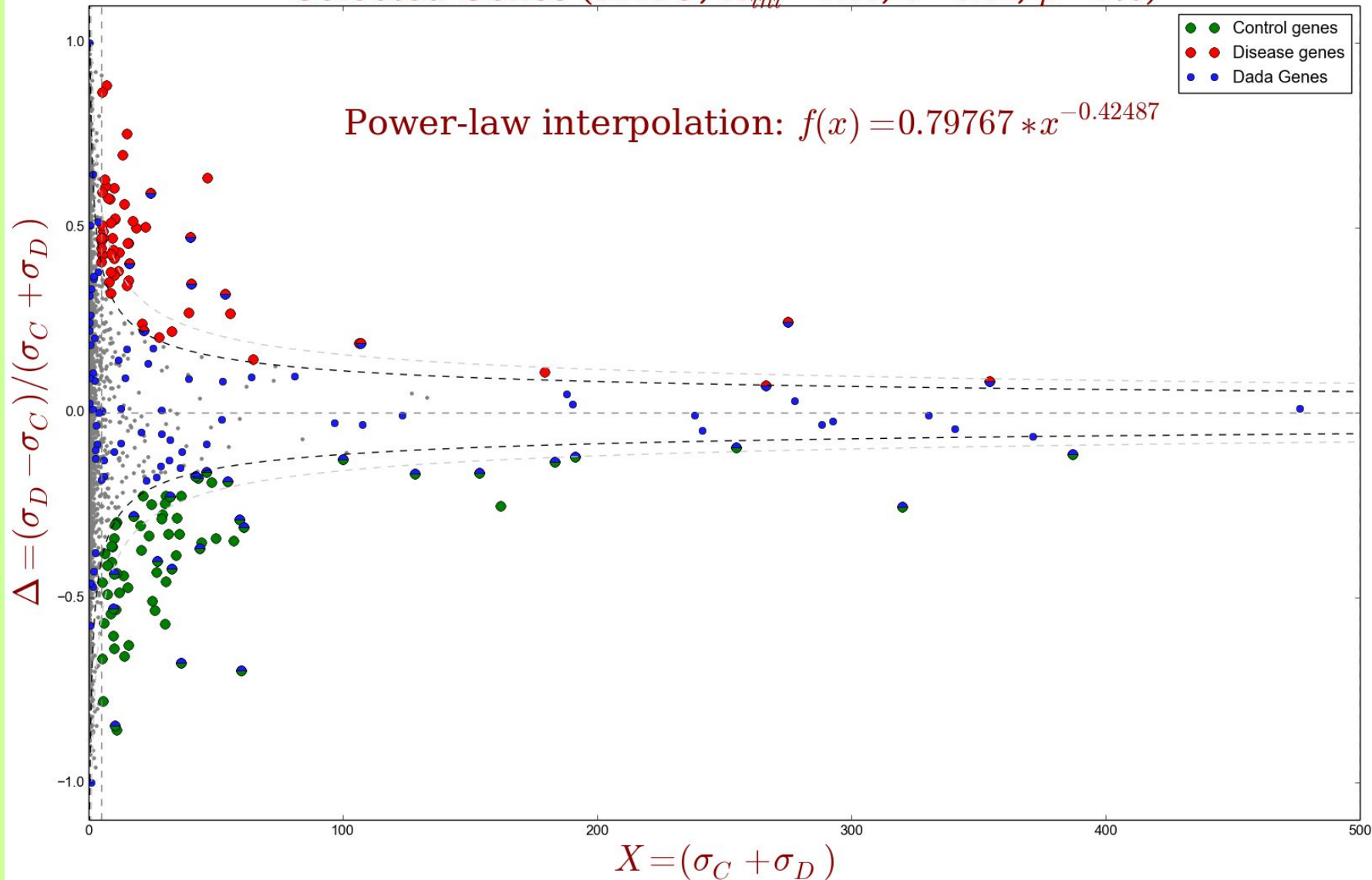


2. Fundamentos conceituais – Método NERI



2. Fundamentos conceituais – Método NERI

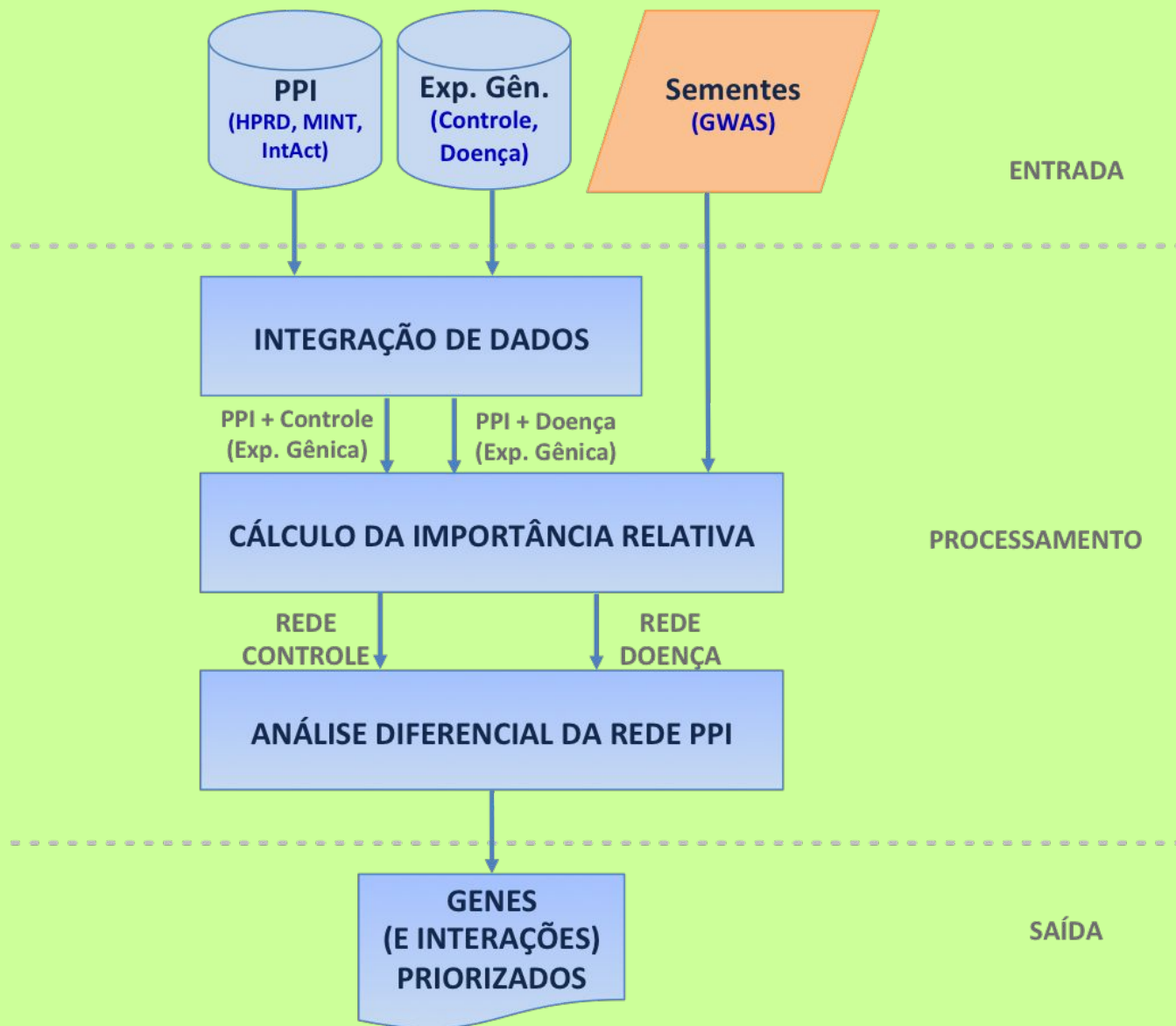
Selected Genes (KATO, $X_{ini}=5.00$, $\epsilon=0.05$, $p=5\%$)



2. Fundamentos conceituais – Avaliação de Classificadores

- Leave-one-out Cross-validation
- Repeated K-Fold Cross-validation

3. Metodologia - Geral



3. Metodologia – Estratégia utilizada para análise de robustez

- Remoção de um único gene semente ($LOO - 30x$)
- Remoção de vários genes sementes (Rep. Kfold – $50x \{3,6,9,12\}$)
- Estudo do impacto causado no resultado (escores X e Δ') em relação ao resultado original

3. Metodologia – Estratégia utilizada para análise de robustez

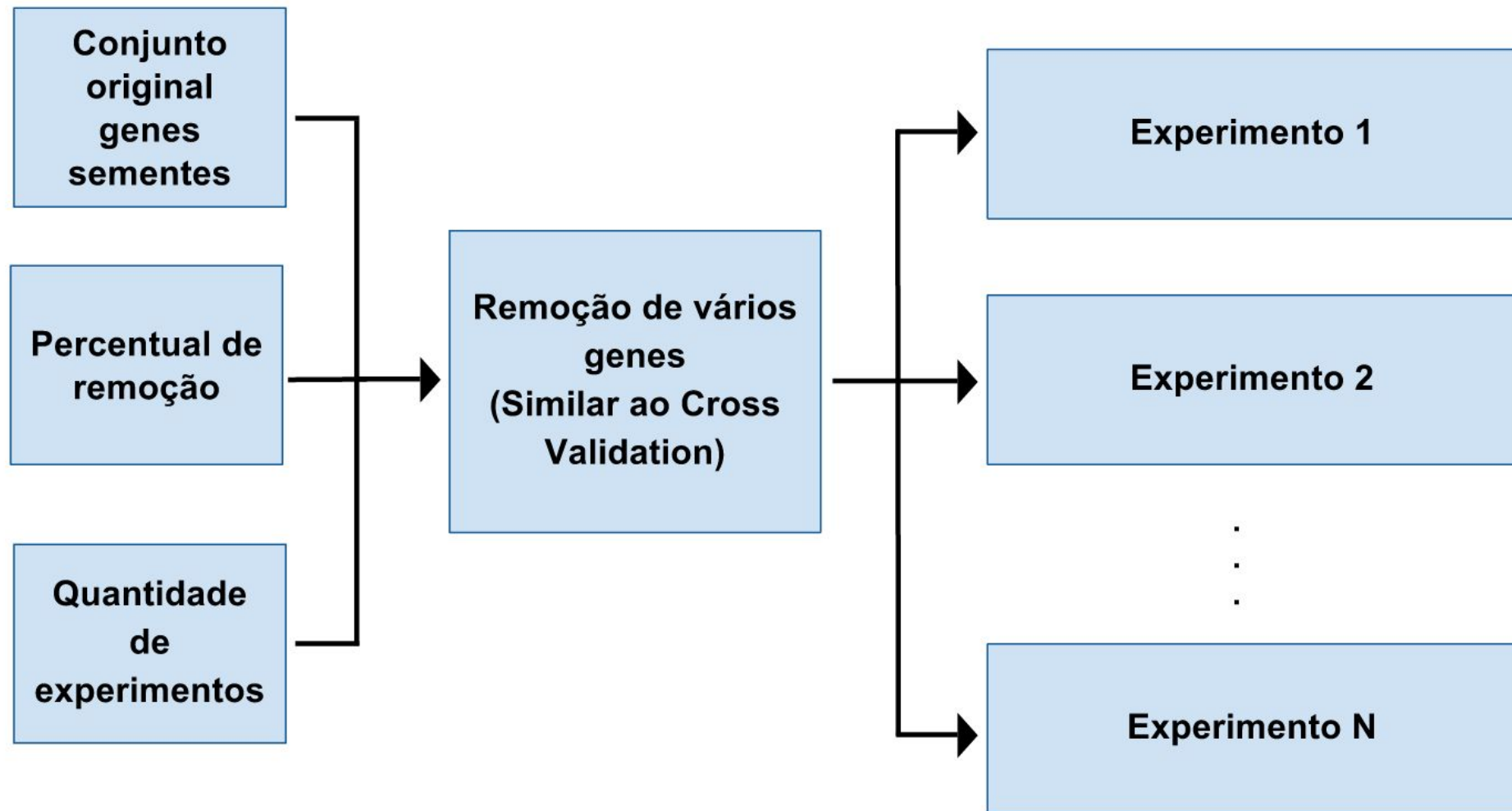
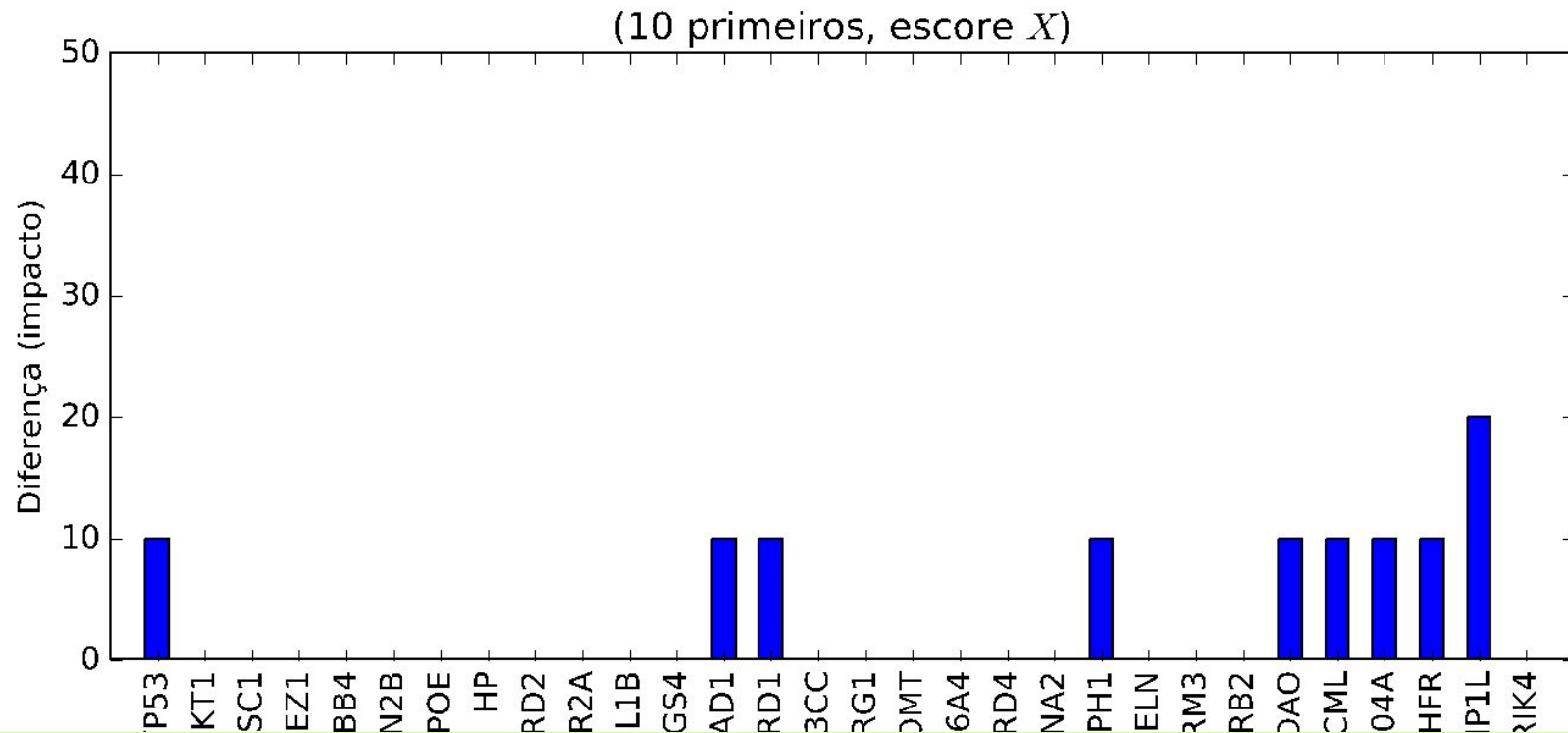


Tabela 4.1 – Medidas de centralidade dos genes sementes utilizados no experimento

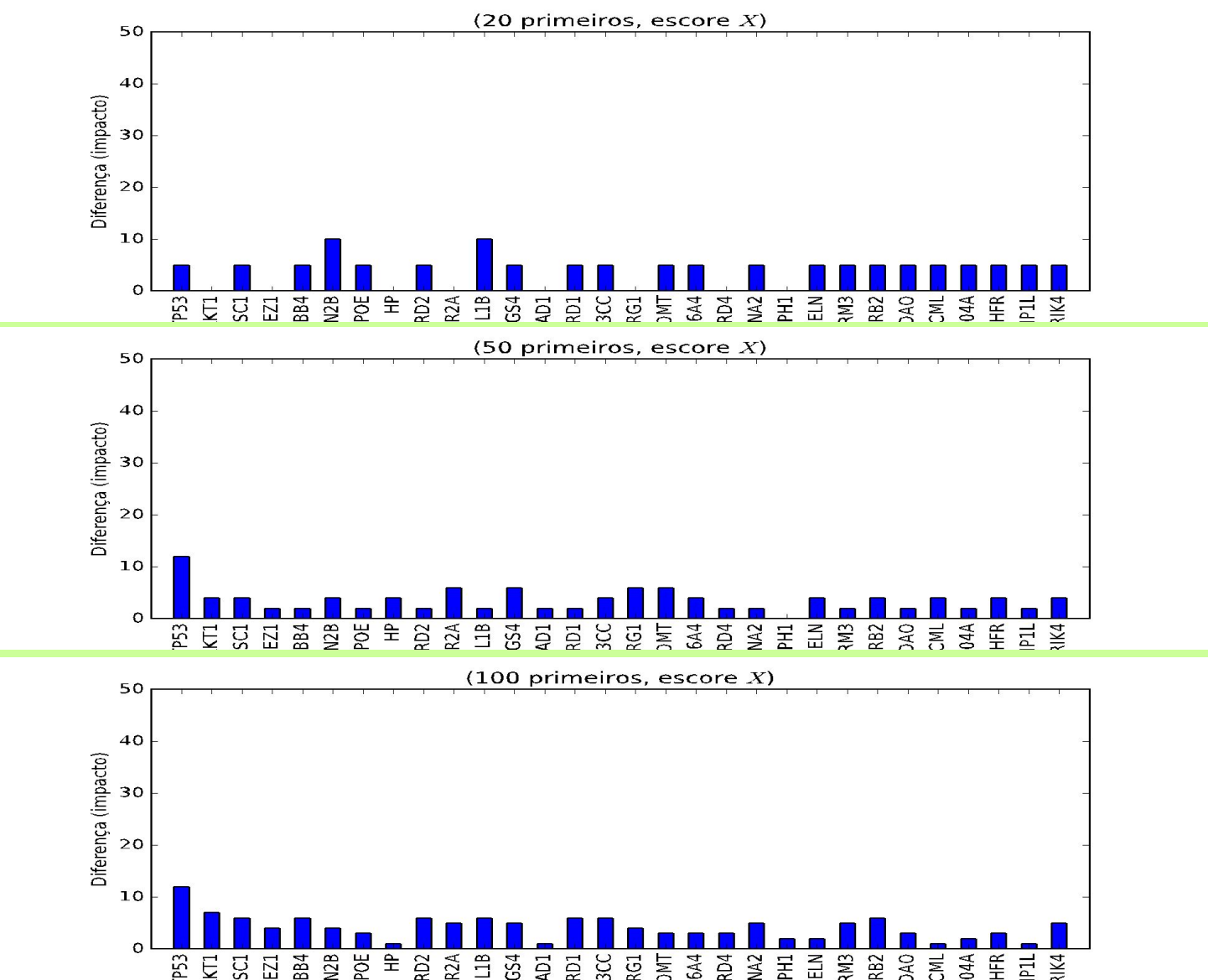
<i>GENE</i>	Degree	Betweenness	Closeness	Clustering	Brokering	Bridgeness
<i>TP53</i>	333	1017443.366745	0.401408	0.027968	0.034839	135.635614
<i>AKT1</i>	138	260150.217669	0.374562	0.044219	0.014196	206.018081
<i>DISC1</i>	91	131642.895006	0.333142	0.016606	0.009632	130.539882
<i>FEZ1</i>	43	39961.948306	0.315902	0.024363	0.004515	196.253855
<i>ERBB4</i>	40	21172.397911	0.325737	0.144872	0.003682	188.374839
<i>GRIN2B</i>	33	23614.446663	0.317337	0.090909	0.003229	362.909288
<i>APOE</i>	29	19099.496744	0.318884	0.088670	0.002845	346.398103
<i>HP</i>	21	22059.197543	0.324588	0.047619	0.002153	519.171834
<i>DRD2</i>	17	15283.833355	0.301050	0.014706	0.001803	580.045729
<i>HTR2A</i>	16	5847.333317	0.292464	0.008333	0.001708	212.879317
<i>IL1B</i>	10	8530.925938	0.304214	0.066667	0.001005	1381.703074
<i>RGS4</i>	10	2039.400569	0.292253	0.066667	0.001005	463.743523
<i>GAD1</i>	10	7993.469172	0.317012	0.222222	0.000837	888.443883
<i>DRD1</i>	8	9964.681533	0.282170	0.071429	0.000800	827.421927
<i>PPP3CC</i>	8	10512.488952	0.289196	0.035714	0.000830	985.201738
<i>NRG1</i>	7	349.958475	0.272072	0.238095	0.000574	126.388843
<i>COMT</i>	5	3676.951004	0.273329	0.000000	0.000538	1028.239964
<i>SLC6A4</i>	5	565.928439	0.283911	0.000000	0.000538	197.019715
<i>DRD4</i>	5	2814.994398	0.298228	0.000000	0.000538	1209.842385
<i>PLXNA2</i>	4	9316.180826	0.264610	0.000000	0.000431	1746.023442
<i>TPH1</i>	4	574.162877	0.312481	0.333333	0.000287	7943.713227
<i>RELN</i>	4	43.733810	0.240987	0.333333	0.000287	13.904311
<i>GRM3</i>	4	248.812705	0.284468	0.000000	0.000431	597.497122
<i>GABRB2</i>	4	555.585080	0.261145	0.000000	0.000431	288.445351
<i>DAO</i>	3	768.669783	0.283496	0.000000	0.000323	676.604177
<i>OPCML</i>	1	0.000000	0.253044	0.000000	0.000108	0.000000
<i>ZNF804A</i>	1	0.000000	0.258773	0.000000	0.000108	0.000000
<i>MTHFR</i>	1	0.000000	0.238457	0.000000	0.000108	0.000000
<i>RPGRIP1L</i>	1	0.000000	0.270930	0.000000	0.000108	0.000000
<i>GRIK4</i>	1	0.000000	0.229498	0.000000	0.000108	0.000000

4. Resultados – Remoção de um único gene - *Score X*

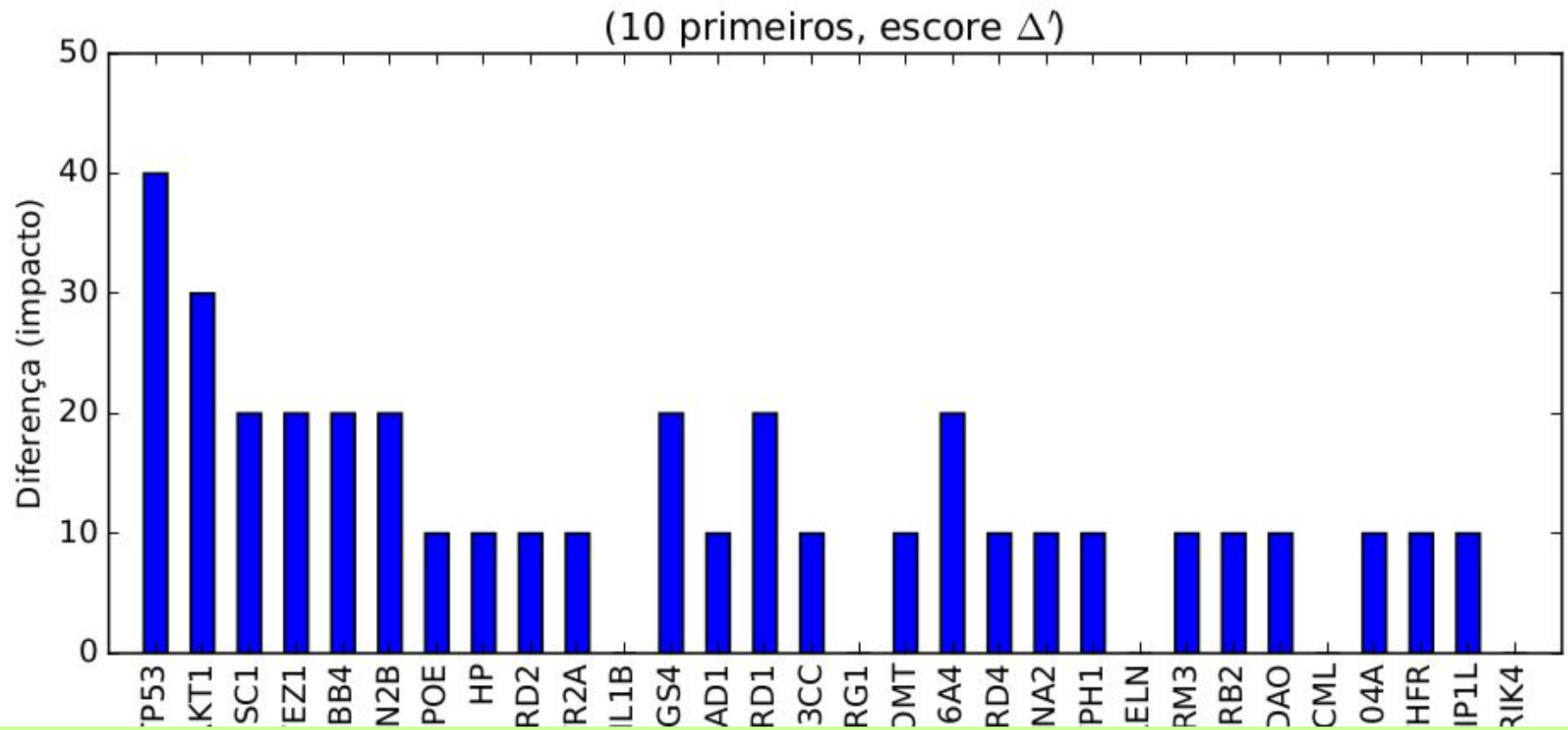
Diferença(%) ou Impacto = (100 – intersecção%)



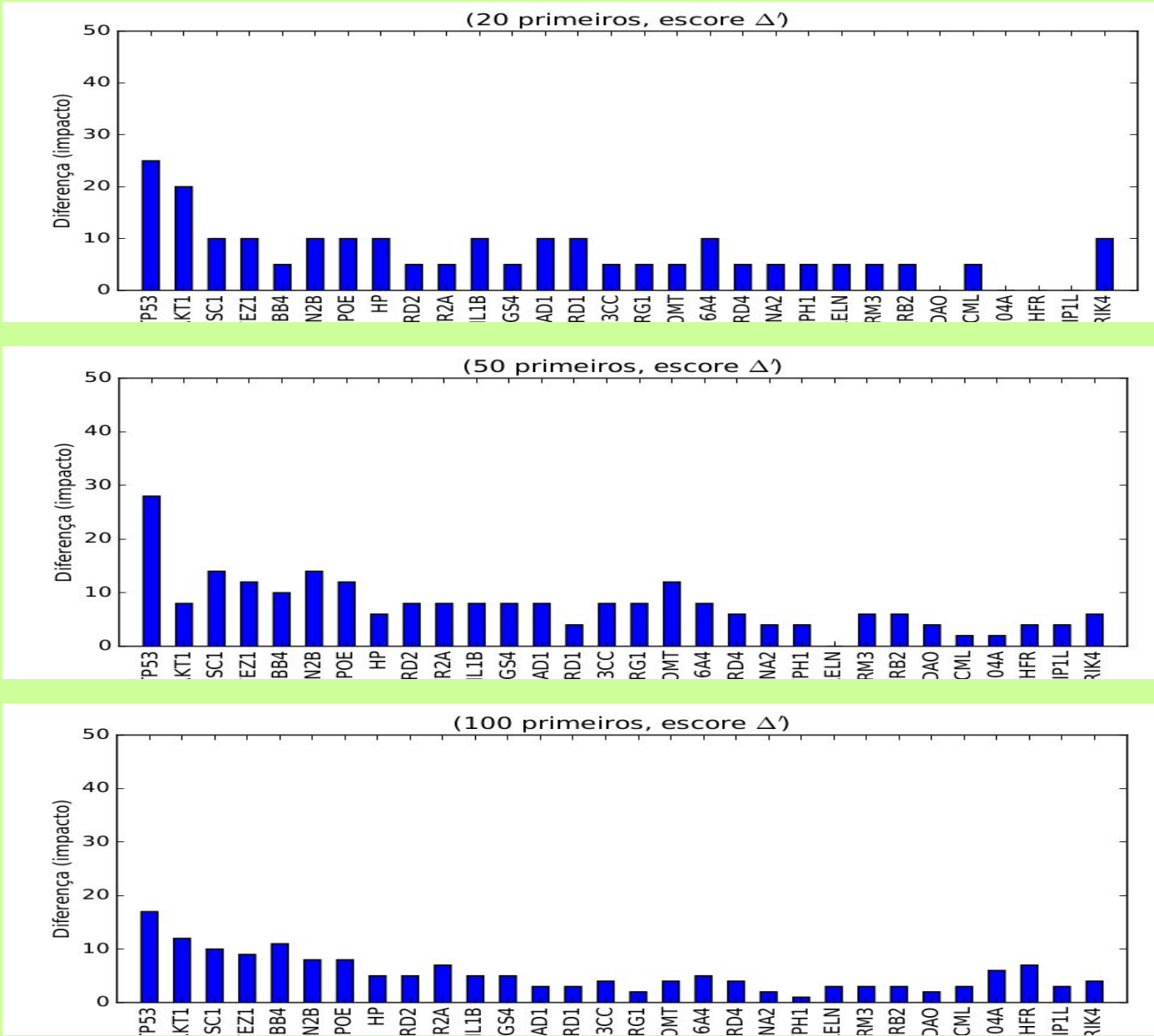
4. Resultados – Remoção de um único gene - *Score X*



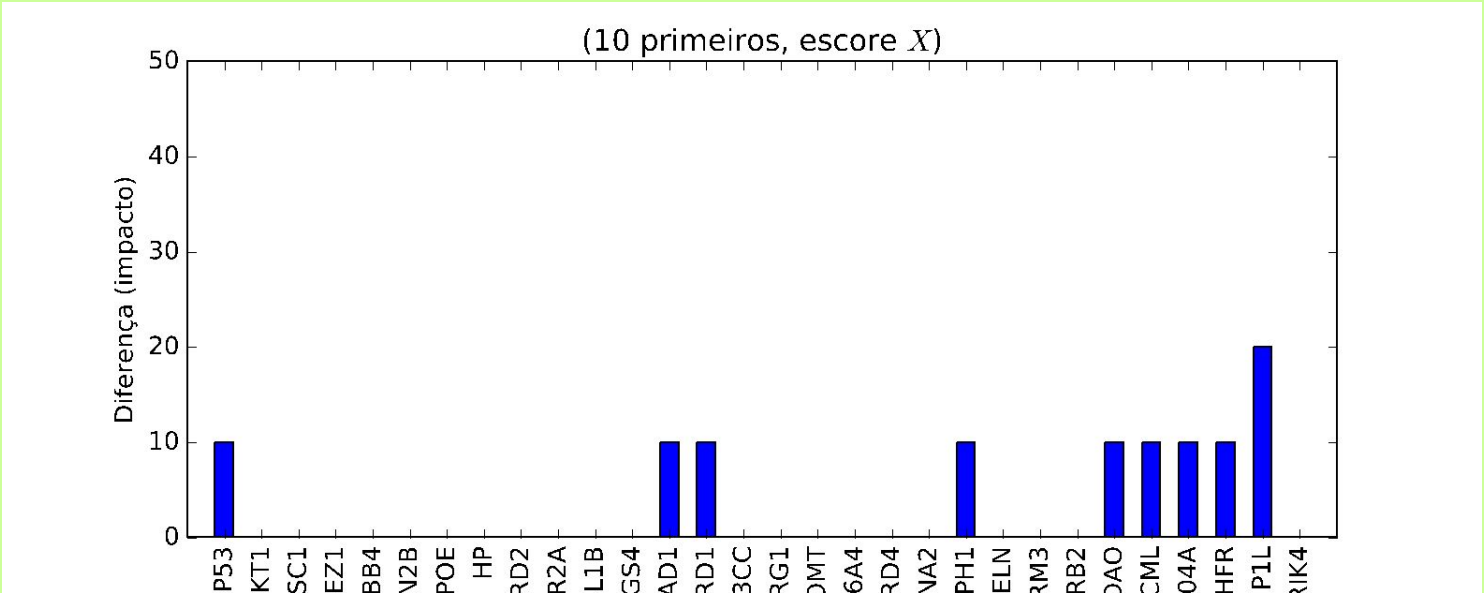
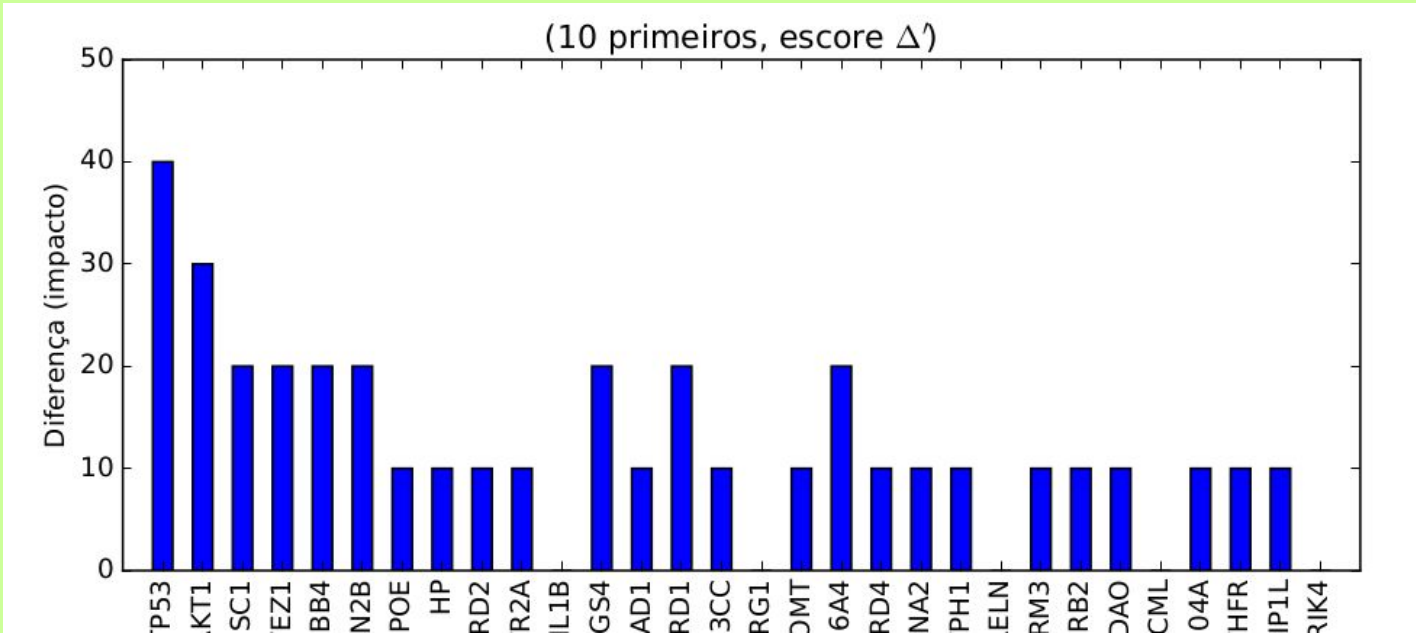
4. Resultados – Remoção de um único gene - $\text{Score } \Delta'$



4. Resultados – Remoção de um único gene - *Score Δ'*

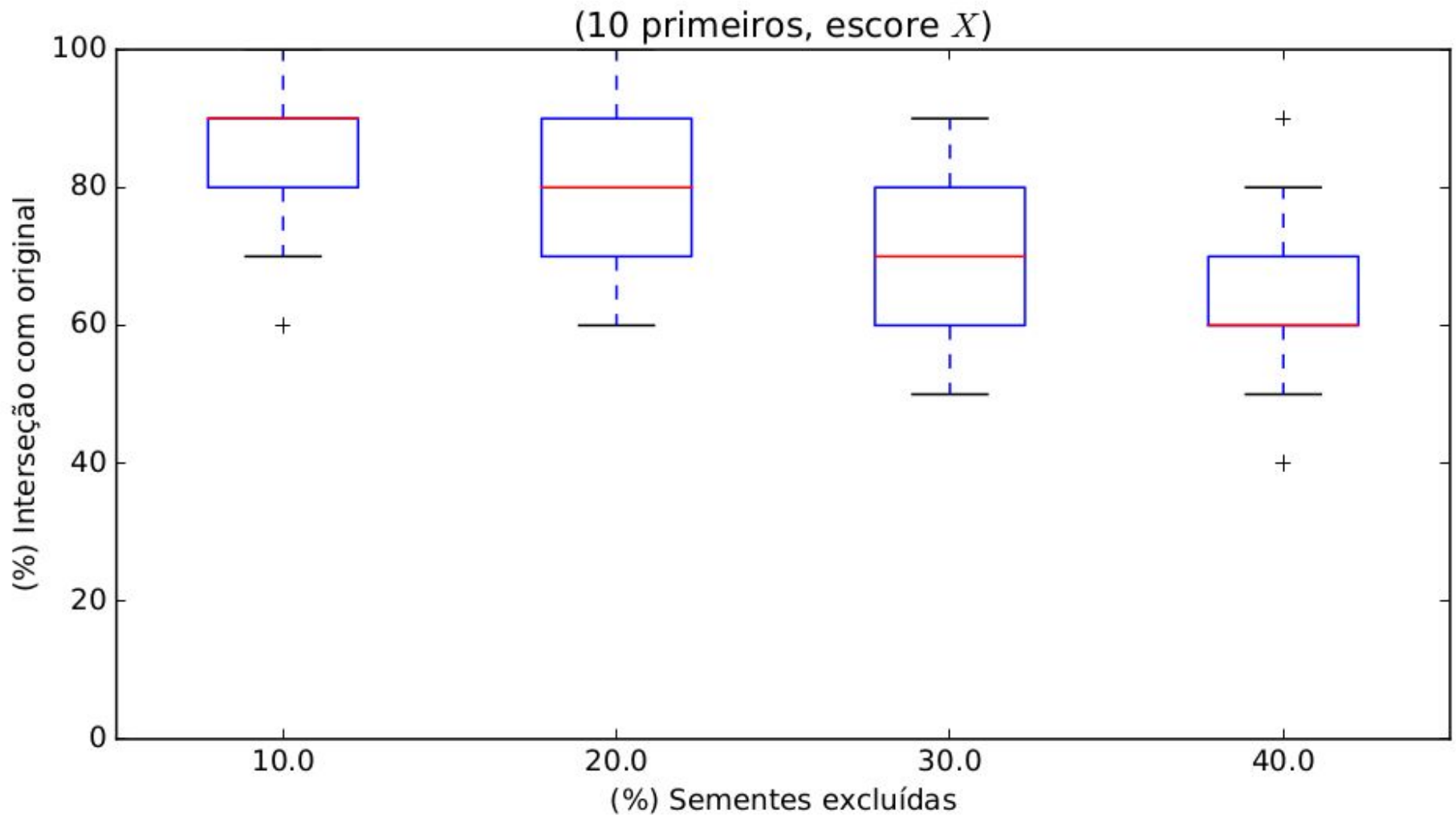


4. Resultados – Remoção de um único gene – $Score \Delta'$ x $Score X$

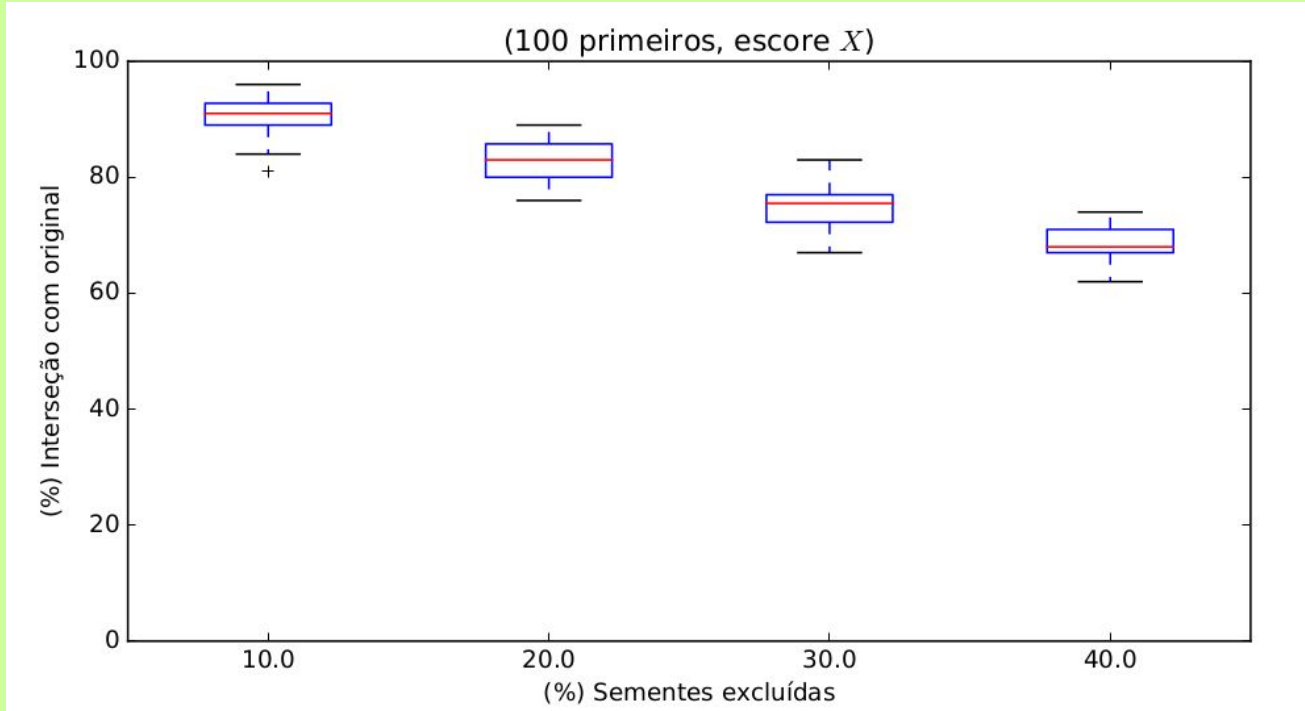
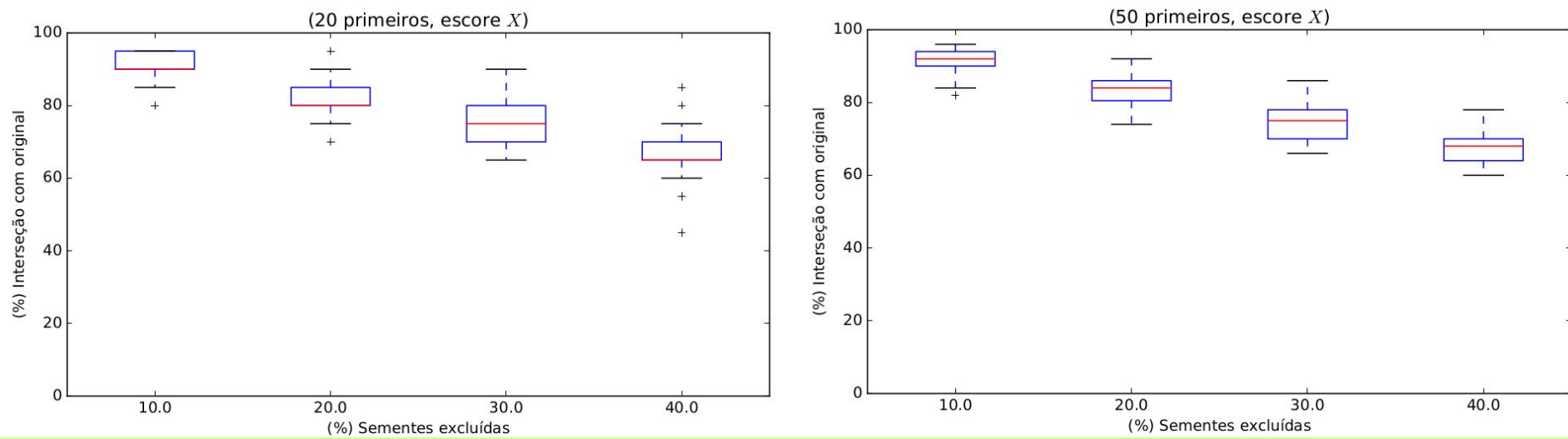


4. Resultados – Remoção de mais de um gene - *Score X*

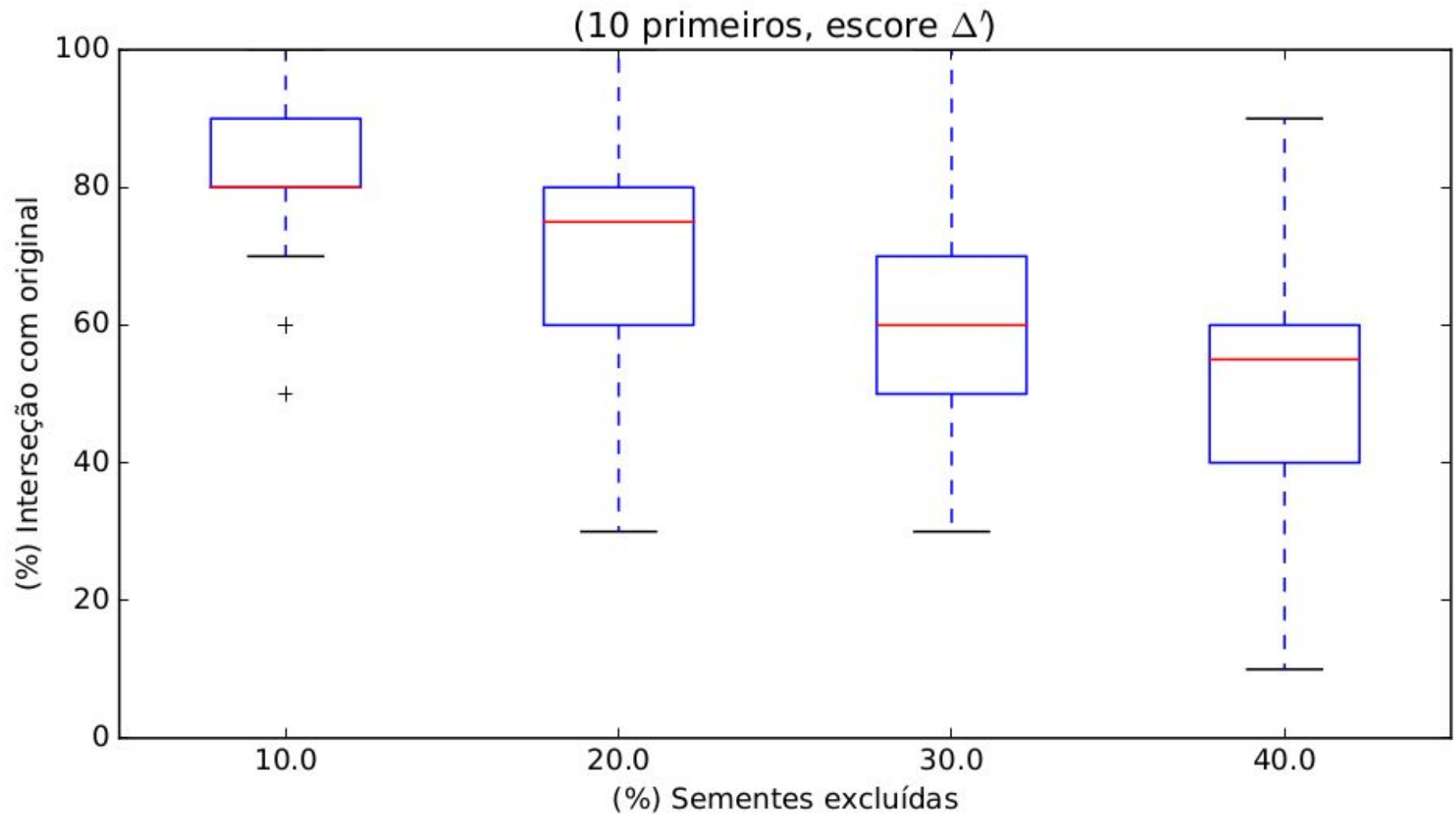
Intersecção = Overlap



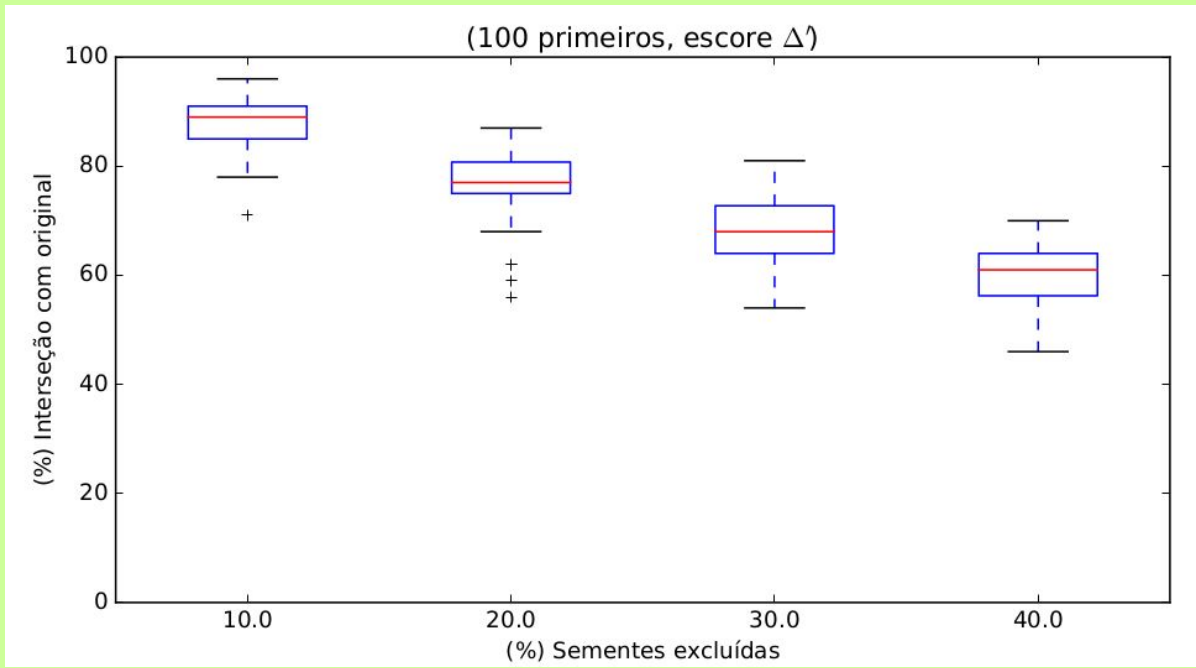
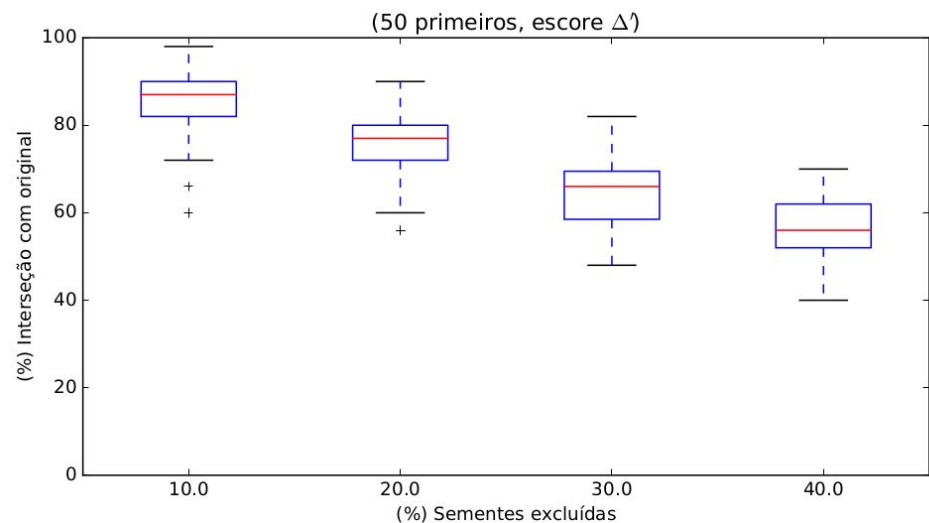
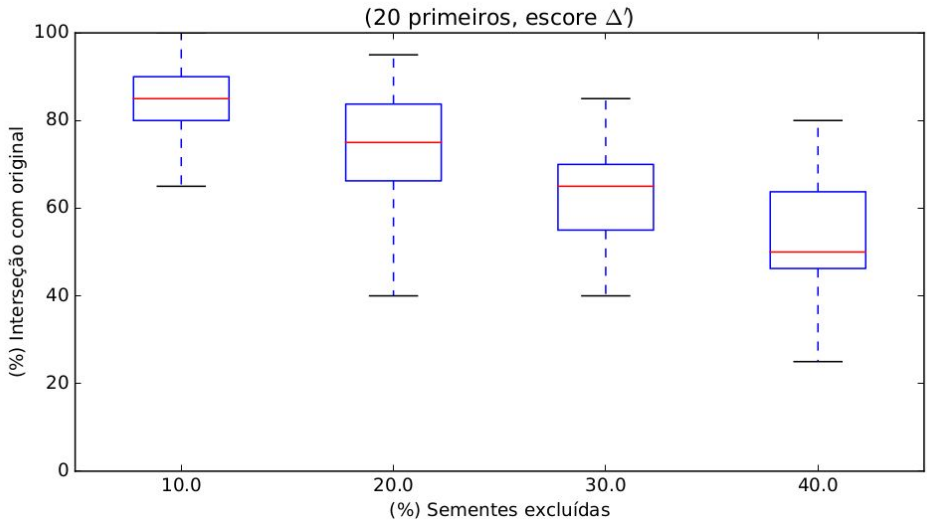
4. Resultados – Remoção de mais de um gene - *Score X*



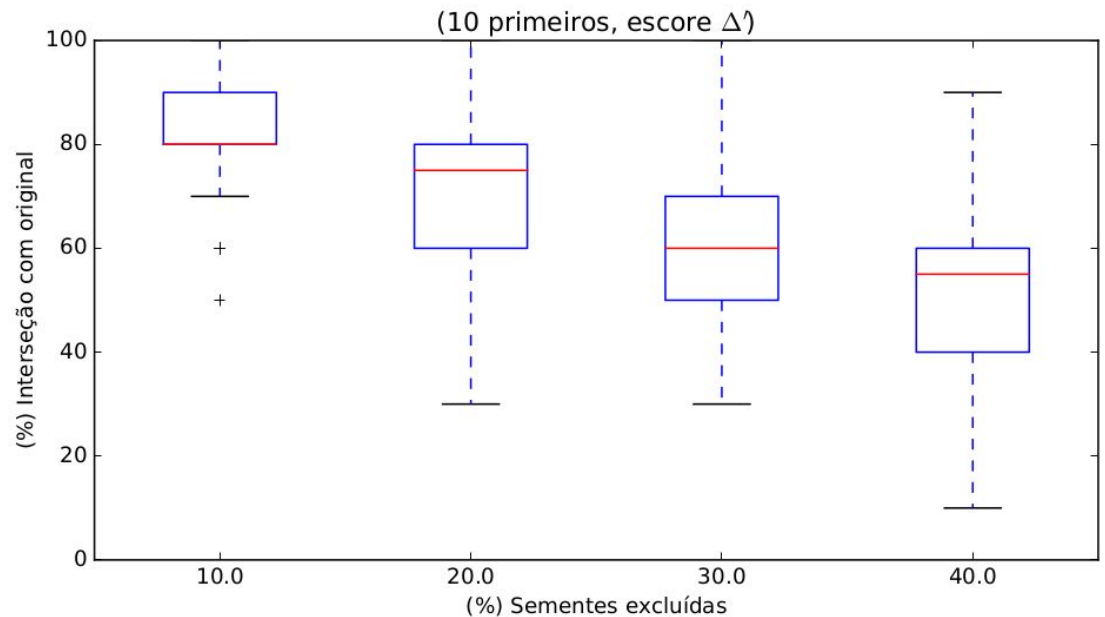
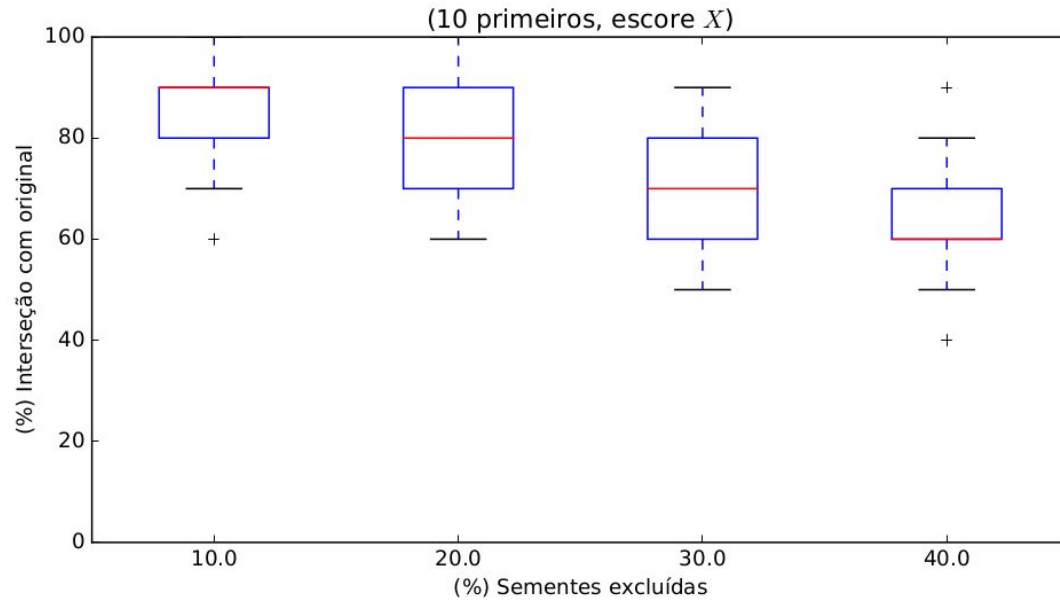
4. Resultados – Remoção de mais de um gene - $\text{Score } \Delta'$



4. Resultados – Remoção de mais de um gene - *Score Δ'*



4. Resultados – Remoção de mais de um gene – $\text{Score } \Delta'$ x $\text{Score } X$



5. Conclusão

- O método apresentou-se robusto em relação aos genes sementes
- Quanto maior a lista comparada, menor é a variância das interseções com o experimento original
- A métrica X é mais robusta que a métrica Δ'
- O Grau do gene semente não está diretamente relacionado com o impacto

6. Trabalhos futuros

1. Analisar a robustez -- em relação as sementes -- de outros métodos, tais como: *Random Walk with Restart* e comparar com os resultados obtidos neste trabalho
2. Pesquisar como integrar novas fontes de dados ao sistema, tais como: dados de epigenética, dados clínicos, etc.
3. Concluir interface gráfica adicionando documentação com tutorial de utilização.
4. Disponibilização do código fonte na web, e possivelmente uma publicação em um *Application Notes*.
5. Criar um serviço web para utilização do método NERI.

Muito obrigado!