

# PONTIFÍCIA UNIVERSIDADE CATÓLICA DE MINAS GERAIS

## Sistemas de Informação – Algoritmos em Grafos

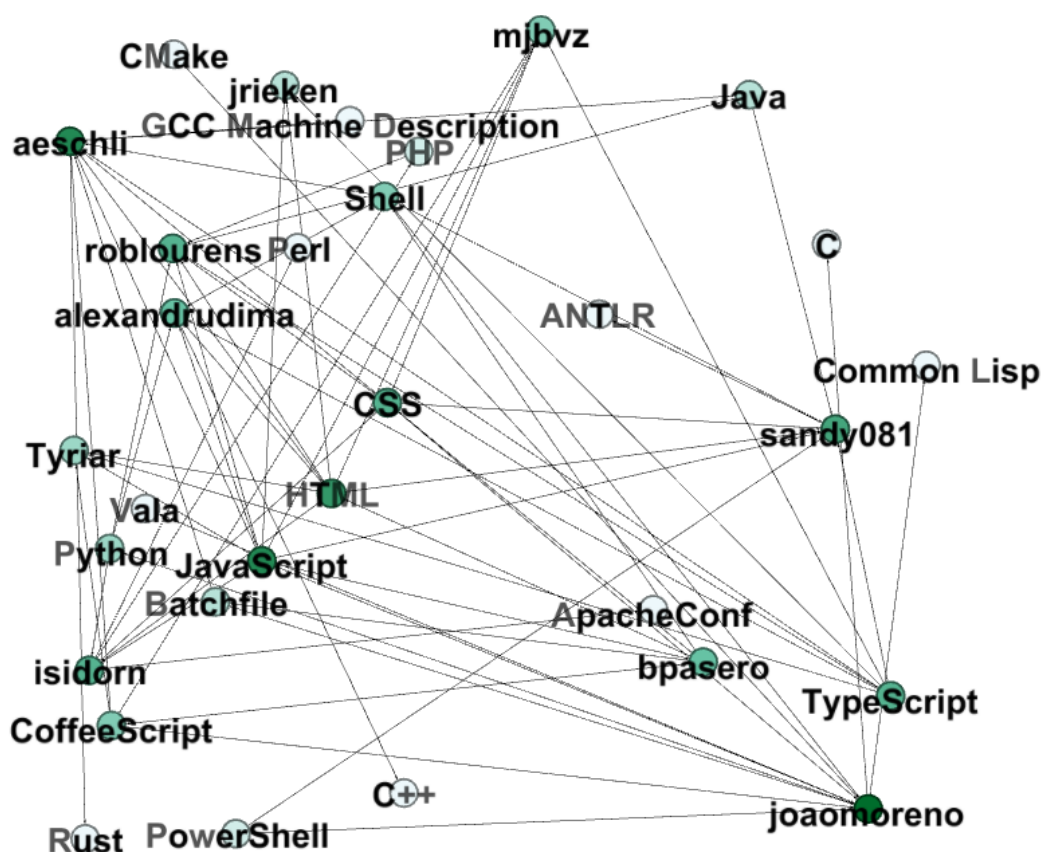
### Trabalho Prático II – Crawler de Rede

**Nome:** Leonardo Willian Ávila Silva

Vinícius Eduardo Santos Gonçalves

*O objetivo deste crawler era verificar a popularidade das linguagens de programação dentre os contribuidores do VSCODE da Microsoft. Para isso, foi feito um pequeno crawler de dados, retirando usuários, e destes usuários seus repositórios, e destes repositórios, a linguagem de programação usada. Neste processo, foi necessário utilizar a biblioteca CSV, Requests, JSON, NetworkX e a API do GitHub. Vale destacar, que os resultados foram reduzidos, devido a um limite de requisições por hora imposto pela API do GitHub, a qual não houve sucesso no uso de Token para autenticação.*

*Segue abaixo os resultados do crawler:*



Grafo gerado pelo Gephi, a partir dos dados coletados no Crawler.

*Note que a cor dos nós foi pintada de verde, com uma coloração cada vez mais escura, para quanto maior for o grau daquele nó. Este é um grafo bipartido(mesmo que não aparentemente, é possível dividir seus nós em dois tipos diferentes: usuário e linguagem de programação), não direcionado, e sem pesos de aresta. Com isso, a coloração dos nós nos ajuda a identificar quais as maiores popularidades.*

*Nota-se, que dentre os nós mais escuros, destaca-se, dentre os de linguagem de programação, os nós JavaScript, HTML e CSS, o que nos permite concluir que as linguagens de programação Web são as mais influentes dentre os usuários analisados.*

*Segue abaixo, alguns resultados de métricas geradas pela biblioteca do NetworkX, a respeito do grafo:*

**Média do grau dos nodos do grafo:** {10: 5.9, 6: 6.25, 7: 6.178571428571429, 5: 7.666666666666667, 3: 8.666666666666666, 9: 6.222222222222222, 11: 3.6363636363636362, 4: 7.75, 1: 9.2, 2: 8.25, 8: 5.5}

**Média de Clustering do Grafo:** 0.0

**Distância média dos nodos:** 2.529032258064516

**Betweenness das arestas:** {('JavaScript', 'bpasero'): 0.02399733657818321, ('JavaScript', 'joaomoreno'): 0.09394148094658987, ('JavaScript', 'jrieken'): 0.03355600769537822, ('JavaScript', 'alexandrudima'): 0.04068652190968353, ('JavaScript', 'isidorn'): 0.0678757793742581, ('JavaScript', 'aeschli'): 0.0490169607097618, ('JavaScript', 'Tyriar'): 0.026917377332838954, ('JavaScript', 'sandy081'): 0.04110205444096558, ('JavaScript', 'mjbvz'): 0.021839988926777826, ('JavaScript', 'roblourens'): 0.030662278418821635, ('bpasero', 'CSS'): 0.01958603370577976, ('bpasero', 'CoffeeScript'): 0.01566210999437976, ('bpasero', 'Shell'): 0.017182199506829202, ('bpasero', 'Batchfile'): 0.01838573275280728, ('bpasero', 'HTML'): 0.023562049869259128, ('CSS', 'joaomoreno'): 0.06501636721057509, ('CSS', 'isidorn'): 0.054139207684650345, ('CSS', 'aeschli'): 0.04070049433826696, ('CSS', 'sandy081'): 0.034346183633861375, ('CSS', 'mjbvz'): 0.019801239016760563, ('CSS', 'roblourens'): 0.026581138794513126, ('CoffeeScript', 'joaomoreno'): 0.04613576325047103, ('CoffeeScript', 'aeschli'): 0.032201386972935696, ('CoffeeScript', 'Tyriar'): 0.019039113123929585, ('CoffeeScript', 'mjbvz'): 0.01740324270622555, ('Shell', 'joaomoreno'): 0.04768203138841998, ('Shell', 'alexandrudima'): 0.03102737956802844, ('Shell', 'aeschli'): 0.03379351768770698, ('Shell', 'sandy081'): 0.02924810373402376, ('Batchfile', 'joaomoreno'): 0.037306892689998095, ('Batchfile', 'aeschli'): 0.035697697616954294, ('HTML', 'jrieken'): 0.02067135562666544, ('HTML', 'alexandrudima'): 0.03254175788616533, ('HTML', 'isidorn'): 0.05323011929339223, ('HTML', 'aeschli'): 0.04229905016603171, ('HTML', 'Tyriar'): 0.020377424063186014, ('HTML', 'sandy081'): 0.03588311629099296, ('HTML', 'mjbvz'): 0.017878978331495234, ('HTML', 'roblourens'): 0.023768716896505948, ('joaomoreno', 'Python'): 0.05820883005483543, ('joaomoreno', 'C'): 0.06451612903225806, ('joaomoreno', 'Vala'): 0.06451612903225806, ('joaomoreno', 'CMake'): 0.06451612903225806, ('joaomoreno', 'PowerShell'): 0.041610211977286496, ('joaomoreno', 'Common Lisp'): 0.06451612903225806, ('Python', 'alexandrudima'): 0.030452163124740507, ('Python', 'isidorn'): 0.03737426318014228, ('Python',

'roblourens'): 0.026965788037321765, ('PowerShell', 'sandy081'): 0.03877805011286658, ('TypeScript', 'jrieken'): 0.015334136521992316, ('TypeScript', 'alexandrudima'): 0.026306181613881596, ('TypeScript', 'aeschli'): 0.0344222891517159, ('TypeScript', 'Tyriar'): 0.015683691756272402, ('TypeScript', 'sandy081'): 0.028341172345267025, ('TypeScript', 'mjbvz'): 0.015620678272905376, ('TypeScript', 'roblourens'): 0.023344746879376856, ('alexandrudima', 'C++'): 0.06451612903225806, ('isidorn', 'PHP'): 0.03165793883535819, ('isidorn', 'ApacheConf'): 0.06451612903225806, ('isidorn', 'Perl'): 0.06451612903225806, ('PHP', 'roblourens'): 0.037226295210166176, ('aeschli', 'Rust'): 0.06451612903225806, ('aeschli', 'GCC Machine Description'): 0.06451612903225806, ('aeschli', 'Java'): 0.0321496376453682, ('Java', 'sandy081'): 0.025364277299761165, ('Java', 'roblourens'): 0.02428642558860775, ('sandy081', 'ANTLR'): 0.06451612903225806}

**Betweenness dos nodos:**{'JavaScript': 0.1951287823505348, 'bpasero': 0.028786885079730835, 'CSS': 0.10457397579166591, 'CoffeeScript': 0.035236036163555, 'Shell': 0.05046431359371135, 'Batchfile': 0.014363793359526731, 'HTML': 0.10994120036438815, 'joaomoreno': 0.3118439466045423, 'Python': 0.04729366166048687, 'C': 0.0, 'Vala': 0.0, 'CMake': 0.0, 'PowerShell': 0.008483381461978372, 'Common Lisp': 0.0, 'TypeScript': 0.050528272289375106, 'jrieken': 0.0026966637097433767, 'alexandrudima': 0.08605920908926691, 'C++': 0.0, 'isidorn': 0.16504476826554887, 'PHP': 0.0023346768174354383, 'ApacheConf': 0.0, 'Perl': 0.0, 'aeschli': 0.19497779418881012, 'Rust': 0.0, 'GCC Machine Description': 0.0, 'Java': 0.009238113043893976, 'Tyriar': 0.009354237837293725, 'sandy081': 0.12456813264810157, 'ANTLR': 0.0, 'mjbvz': 0.01498048180826036, 'roblourens': 0.06858443249283987}

**Observação final: OS RESULTADOS OBTIDOS FORAM UM RECORTE DO ORIGINAL, UMA VEZ QUE O ALGORITMO ERA PARADO EM DETERMINADO MOMENTO DEVIDO AO LIMITE DE REQUISIÇÕES DA API, PORÉM, SÃO DADOS CONCRETOS, PASSÍVEIS DE CONFERENCIA NOS RESPECTIVOS REPOSITÓRIOS.**