

Projeto 4  
Teoria e Aplicação de Grafos, Turma A, 1/2017  
Prof. Dúbio

Uma das possibilidades em análise de grafos é evidenciar estruturas anômalas, ou anomalias em subconjuntos de elementos representados em um grafo. Há vários trabalhos nesse sentido, e um deles é o de (Liu et al 2015) (ps. cópia fornecida no moodle da disciplina) que propõe a identificação de anomalias a partir de mapeamento de comunidades/estruturas no grafo  $G$ , extração de atributos caracterizadores dessas estruturas e uso de multivariáveis para apontar anomalias. É fornecido para este trabalho um conjunto de dados públicos no kaggle (Brazil's House of Deputies Reimbursements), os quais são disponibilizados publicamente em (<http://www2.camara.leg.br/transparencia/cota-para-exercicio-da-atividade-parlamentar/dados-abertos-cota-parlamentar>) que relacionam gastos de reembolso da Câmara de Deputados, parlamentar por parlamentar, valores, finalidade, empresas receptoras. A tabela completa possui vários campos extras, mas para este trabalho pede-se usar as 8 primeiras, a saber:

u'deputy\_name': deputy official name  
u'deputy\_state': state from which deputy was elected  
u'political\_party': deputy affiliation party  
u'refund\_description': reason for refund  
u'company\_name': company utilized for expenses  
u'company\_id': company legal id (like a social security number for companies)  
u'refund\_date': unclear whether its the refund request or granted date  
u'refund\_value' : value in R\$ of refund ( 3R\$ ~ 1USD )

O arquivo de dados, dirty\_deputies\_v2.csv, está disponível em  
<https://www.kaggle.com/epattaro/brazils-house-of-deputies-reimbursements>

Neste projeto pede-se usar o arquivo fornecido como entrada, ou reformatar com somente as 8 variáveis pedidas um outro e a partir deste o programa deverá receber esta entrada, montar um grafo  $G$  adequado, implementar o algoritmo de (Liu et al 2015) para detectar estruturas anômalas e fornecer em tela e em arquivo de saída os resultados constando quais seriam essas estruturas anômalas no grafo.

Liu, J.; Bier, E.; Wilson, A.; Honda, T.; Kumar, S.; Gilpin, L.; Guerra-Gomez, J. and Davies, D. (2015). Graph Analysis for Detecting Fraud, Waste, and Abuse in Healthcare Data. *Proceedings of the Twenty-Seventh Conference on Innovative Applications of Artificial Intelligence*. AAAI Press, pp. 3912-3919.

O código deve ser bem documentado, de forma modular com funções para cada tarefa independente, realizado por dois (2) estudantes do curso usando “pair programming”, e entregue via sistema <http://aprender.unb.br> do curso, no prazo estipulado.