



Aula 5 - Análise combinada de uma amostra do Twitter (Pt. 1)

José Devezas <jld@fe.up.pt>

Faculdade de Engenharia da Universidade do Porto

Gestão da Informação em Redes Sociais

Mestrado em Ciência da Informação



Conteúdos

- Introdução
- Tutorial de Gephi (cont.)
 - Separar grafo da comunidade do grafo original
 - Comparar métricas globais e na comunidade
 - Geração de grafos aleatórios e *small-world effect*
- 3ª avaliação



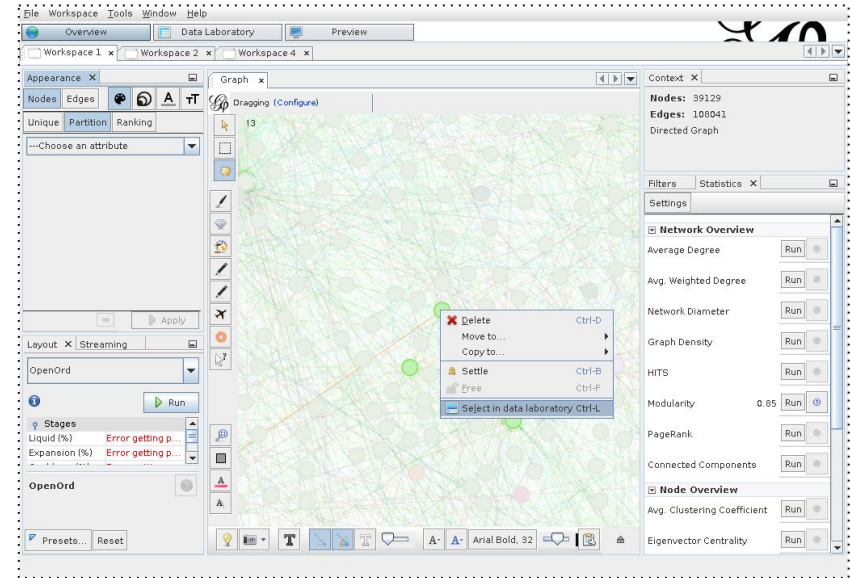
Introdução

- Estas duas aulas serão dedicadas à realização do trabalho de grupo.
- Apresentaremos informação adicional sobre a identificação do *small-world effect*.
- E depois devem formar grupos e começar a trabalhar.

Tutorial de Gephi (cont.)

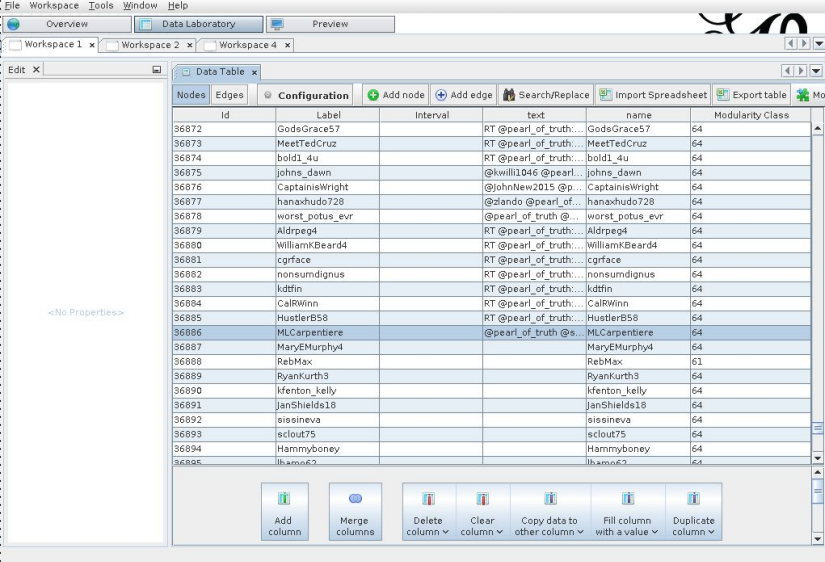
Separar grafo da comunidade do grafo original

Após escolher uma comunidade com base nos indicadores visuais, fazer zoom e clicar com o botão direito num dos nós da comunidade, selecionando “Select in data laboratory”.



Separar grafo da comunidade do grafo original

Visitando o “Data Laboratory”, podemos verificar que o registo correspondente ao nó se encontra selecionado e dentro da área de visualização. Verificamos que a “Modularity Class” para este nó é 64.



The screenshot shows the 'Data Laboratory' application interface. The 'Nodes' tab is active, displaying a table with columns: Id, Label, Interval, text, name, and Modularity Class. The table contains 20 rows of data. The row with Id 36886 is highlighted in blue. The 'Modularity Class' column for all rows shows the value 64. The interface includes a menu bar (File, Workspace, Tools, Window, Help), a toolbar with buttons like 'Add node', 'Add edge', 'Search/Replace', 'Import Spreadsheet', 'Export table', and a bottom toolbar with buttons like 'Add column', 'Merge columns', 'Delete column', 'Clear column', 'Copy data to other column', 'Fill column with a value', and 'Duplicate column'.

Id	Label	Interval	text	name	Modularity Class
36872	GodsGrace57		RT @pearl_of_truth...	GodsGrace57	64
36873	MeetTedCruz		RT @pearl_of_truth...	MeetTedCruz	64
36874	bold1_4u		RT @pearl_of_truth...	bold1_4u	64
36875	johns_dawn		@kwill1046 @pearl...	johns_dawn	64
36876	CaptainisWright		@johnNew2015 @p...	CaptainisWright	64
36877	hanaxhudo728		@zlando @pearl of...	hanaxhudo728	64
36878	worst_potus_evr		@pearl_of_truth @...	worst_potus_evr	64
36879	Aldrpeg4		RT @pearl_of_truth...	Aldrpeg4	64
36880	WilliamkBeard4		RT @pearl_of_truth...	WilliamkBeard4	64
36881	cgrface		RT @pearl_of_truth...	cgrface	64
36882	nonsumdignus		RT @pearl_of_truth...	nonsumdignus	64
36883	ldtfin		RT @pearl_of_truth...	ldtfin	64
36884	CalRWinn		RT @pearl_of_truth...	CalRWinn	64
36885	HuslerB58		RT @pearl_of_truth...	HuslerB58	64
36886	MLCarpentiere		@pearl_of_truth @s...	MLCarpentiere	64
36887	MaryEMurphy4			MaryEMurphy4	64
36888	RebMax			RebMax	61
36889	Ryankurth3			Ryankurth3	64
36890	kfenton_kelly			kfenton_kelly	64
36891	JanShields18			JanShields18	64
36892	sisineva			sisineva	64
36893	sclout75			sclout75	64
36894	Hammyboney			Hammyboney	64
36895	libamof2			libamof2	64

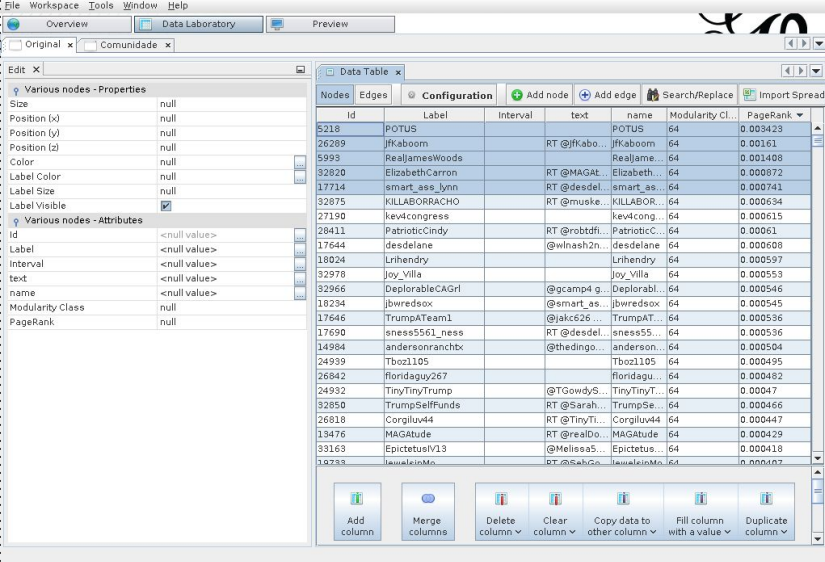
Separar grafo da comunidade do grafo original

Vamos utilizar um filtro de igualdade de atributo para o “Modularity Class” com o valor 64, exportando o resultado para um novo Workspace.

The screenshot displays the Data Laboratory software interface. The central workspace shows a dense network graph with green circular nodes and green edges. On the right side, the 'Filters' panel is open, showing a tree structure under 'Library' with 'Equal' selected. A red circle (1) highlights the 'Equal' filter. Below it, the 'Attributes' list includes 'Modularity Class Integer (Node)', which is also highlighted with a red circle (2). The 'Queries' panel shows 'Equal (Modularity Class)' selected. At the bottom right, the 'Settings' for the selected query are visible, with the 'Value' field set to '64' (highlighted with a red circle 3) and a 'Stop' button (highlighted with a red circle 4). The top of the interface shows tabs for 'Workspace 1', 'Workspace 2', and 'Workspace 4'. The bottom status bar indicates 'Arial Bold, 32'.

Comparar métricas globais e na comunidade

Calculamos o PageRank para o **grafo original** e, mantendo o filtro ligado, ordenamos por PageRank no “Data Laboratory”, selecionando os primeiros 5 elementos.

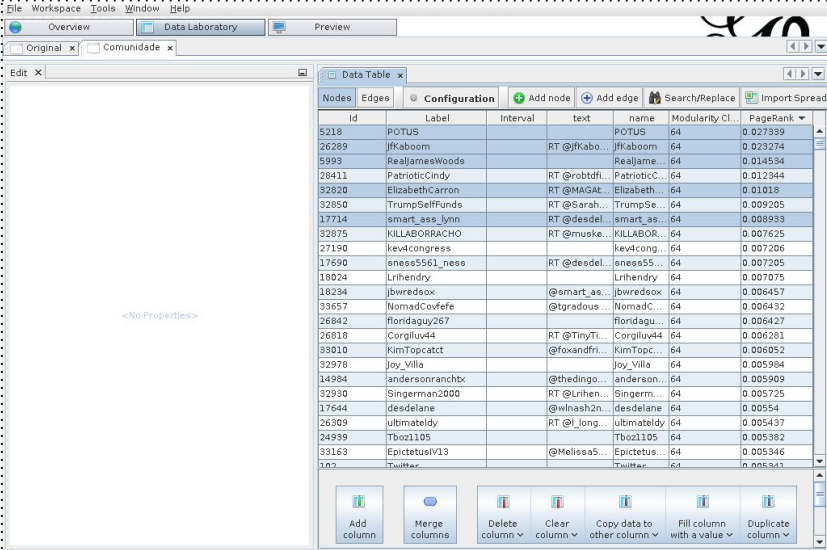


The screenshot shows the 'Data Laboratory' application interface. On the left, the 'Edit' panel shows properties for 'Various nodes' (Size, Position, Color, Label, Label Size, Label Visible) and attributes (Id, Label, Interval, text, name, Modularity Class, PageRank). The main area displays a table of nodes, sorted by PageRank in descending order. The table has columns: Id, Label, Interval, text, name, Modularity Cl., and PageRank. The bottom of the interface features a toolbar with buttons for 'Add column', 'Merge columns', 'Delete column', 'Clear column', 'Copy data to other column', 'Fill column with a value', and 'Duplicate column'.

Id	Label	Interval	text	name	Modularity Cl.	PageRank
5218	POTUS			POTUS	64	0.003429
26289	JKaboom		RT @JKabo...	JKaboom	64	0.00161
5999	RealJamesWoods			Realjame...	64	0.001408
32820	ElizabethCarron		RT @MAGat...	Elizabeth...	64	0.000872
17714	smart_ass_lynn		RT @desdel...	smart_as...	64	0.000741
32875	KILLABORRACHO		RT @muske...	KILLABOR...	64	0.000634
27190	kev4congress			kev4cong...	64	0.000615
28411	PatrioticCindy		RT @robtdf...	PatrioticC...	64	0.00061
17644	desdelane		@wlnash2n...	desdelane	64	0.000608
18024	Lrihendry			Lrihendry	64	0.000597
32978	Joy_Villa			Joy_Villa	64	0.000553
32966	DeplorableCAGrl		@gcamp4 g...	Deplorabl...	64	0.000546
18234	jbwredsox		@smart_as...	jbwredsox	64	0.000545
17646	TrumpATeam1		@jakc26	TrumpAT...	64	0.000536
17690	sness5561_ness		RT @desdel...	sness55...	64	0.000536
14984	andersonranchbx		@thedingo...	anderson...	64	0.000504
24939	Tboz1105			Tboz1105	64	0.000495
26842	floridaguy267			floridagu...	64	0.000482
24932	TinyTinyTrump		@TGowdyS...	TinyTinyT...	64	0.00047
32850	TrumpSelfFunds		RT @Sarah...	TrumpSe...	64	0.000466
26818	Corgiluv44		RT @TinyTi...	Corgiluv44	64	0.000447
13476	MAGAtude		RT @realDo...	MAGAtude	64	0.000429
33163	EpictetusV13		@Melisea5...	Epictetus...	64	0.000418
16733	laualacinMo...		RT @SahGo...	laualacinM...	64	0.000407

Comparar métricas globais e na comunidade

Calculamos o PageRank para o grafo da comunidade e ordenamos por PageRank no “Data Laboratory”, mantendo a seleção do grafo original. Notamos a presença da @PatrioticCindy e do @TrumpSelfFunds como contas importantes na comunidade.

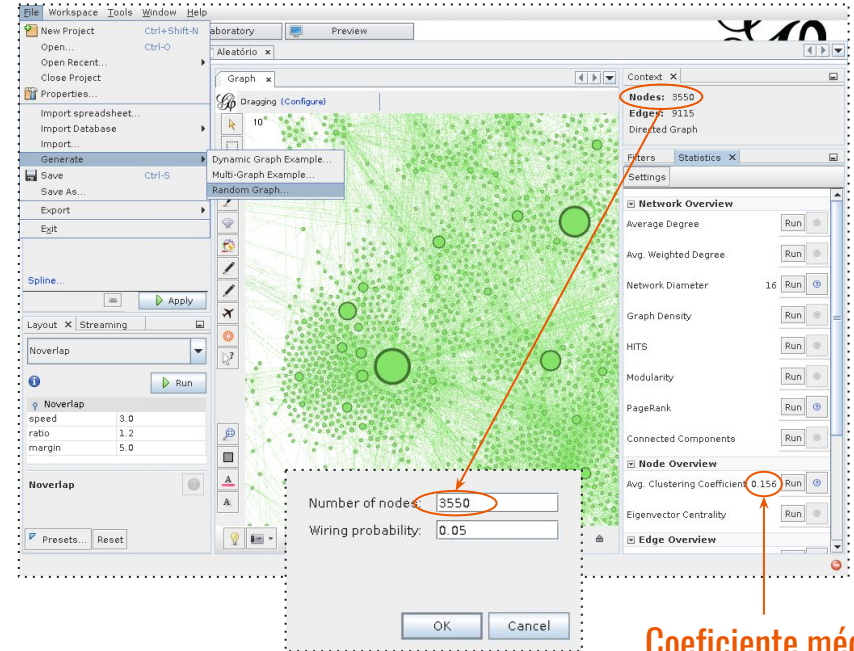


The screenshot shows the Data Laboratory interface with a table of nodes. The table has columns for Id, Label, Interval, text, name, Modularity Cl., and PageRank. The nodes are sorted by PageRank in descending order. The interface includes a menu bar (File, Workspace, Tools, Window, Help), a toolbar (Overview, Data Laboratory, Preview), and a sidebar (Nodes, Edges, Configuration, Add node, Add edge, Search/Replace, Import Spread).

Id	Label	Interval	text	name	Modularity Cl.	PageRank
5218	POTUS			POTUS	64	0.027399
26289	JKaboom		RT @JKabo...	JKaboom	64	0.023274
5999	RealJamesWoode			RealJame...	64	0.014534
28411	PatrioticCindy		RT @robtfd...	PatrioticC...	64	0.012344
32820	ElizabethCarron		RT @MAGat...	Elizabeth...	64	0.01018
32850	TrumpSelfFunds		RT @Sarah...	TrumpSe...	64	0.009205
17714	smart_ass_lynn		RT @desdel...	smart_as...	64	0.008933
32875	KILLABORRACHO		RT @muske...	KILLABOR...	64	0.007625
27190	kev4congress			kev4cong...	64	0.007206
17690	snese5561_ness		RT @desdel...	snese55...	64	0.007205
18024	Lrihendry			Lrihendry	64	0.007075
18234	jbwredsox		@smart_as...	jbwredsox	64	0.006457
33657	NormadCovefe			NormadC...	64	0.006432
26842	floridaguy267			floridagu...	64	0.006427
26818	Corgilu44		RT @TinyTi...	Corgilu44	64	0.006281
33010	KimTopcatt			KimTopc...	64	0.006052
32978	Joy Villa			Joy Villa	64	0.005984
14984	andersonranchb		@thedingo...	anderson...	64	0.005909
32930	Singerman2000		RT @Lrihen...	Singerm...	64	0.005725
17644	desdelane		@wlnash2h...	desdelane	64	0.00554
26309	ultimatelyd		RT @_long...	ultimatelyd	64	0.005437
24939	Tboz1105			Tboz1105	64	0.005382
33163	EpictetusIV13		@Melisea5...	Epictetus...	64	0.005346
183	Twitter			Twitter	64	0.005343

Geração de grafos aleatórios e *small-world effect*

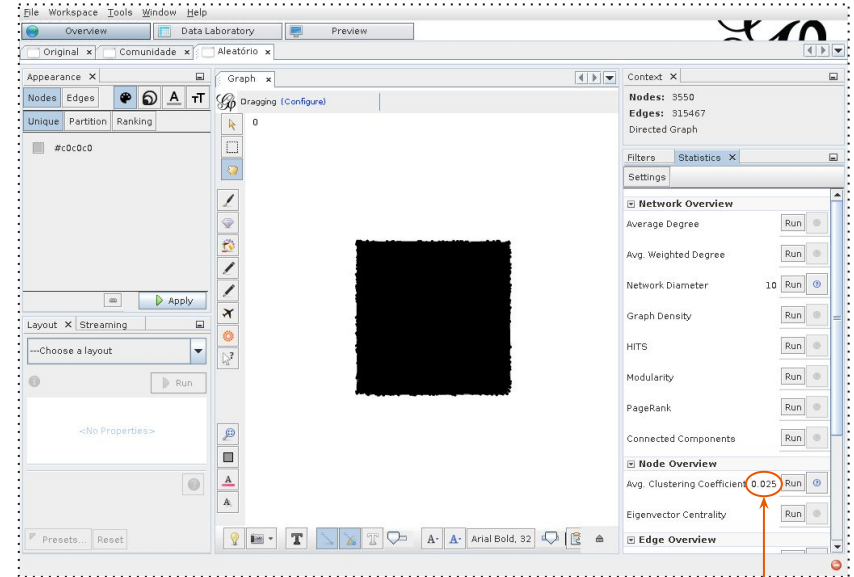
O coeficiente médio de agrupamento no grafo da comunidade é de **0,156**. Vamos gerar um grafo aleatório com o mesmo número de nós do grafo da comunidade para podermos verificar se existe *small-world effect*.



Coeficiente médio de agrupamento

Geração de grafos aleatórios e *small-world effect*

Conforme podemos verificar, quando geramos um grafo com arestas aleatórios com o mesmo número de nós que o grafo da comunidade, o coeficiente médio de agrupamento é menor, sendo apenas de **0,025**. Verifica-se o *small-world effect*.



Coeficiente médio de agrupamento



3ª avaliação

- Em grupo:
 - Devem criar grupos entre duas a quatro pessoas (discutível).
 - O objetivo do trabalho é voltar a pegar no conjunto de dados do Twitter, estudado anteriormente, agora fazendo uma análise combinada (grafo social e texto).
 - Repetir a análise da rede e selecionar uma comunidade diferente para estudar, sendo que, desta vez, a informação deverá ser também analisada textualmente.
 - Poderão exportar a informação textual através da funcionalidade de exportação do “Data Laboratory”, gerando vários ficheiros CSV, por exemplo por comunidade
 - Tirando partido do KNIME, deverão adaptar os Workflows partilhados nas aulas anteriores de forma a permitir tratar os ficheiros CSV gerados.
- Este trabalho será desenvolvido na segunda parte desta aula e durante a próxima aula.
- Deverão entregar um relatório sobre a análise combinada antes da última aula desta componente, onde deverão apresentar o vosso trabalho.